

**Erkki Mäkinen (toim.)**

**Tietojenkäsittelytieteellisiä tutkielmia  
Kesä 2015**



INFORMAATIOTIETEIDEN YKSIKKÖ  
TAMPEREEN YLIOPISTO

INFORMAATIOTIETEIDEN YKSIKÖN RAPORTTEJA 39/2015

TAMPERE 2015

TAMPEREEN YLIOPISTO  
INFORMAATIOTIETEIDEN YKSIKKÖ  
INFORMAATIOTIETEIDEN YKSIKÖN RAPORTTEJA 39/2015  
KESÄKUU 2015

**Erkki Mäkinen (toim.)**

**Tietojenkäsittelytieteellisiä tutkielmia  
Kesä 2015**

INFORMAATIOTIETEIDEN YKSIKKÖ  
33014 TAMPEREEN YLIOPISTO

ISBN 978-951-44-9856-5 (pdf)

ISSN-L 1799-8158  
ISSN 1799-8158

## Sisällysluettelo

Mobiiliavusteisen kielen oppimisen osa-alueet ja tehokkuus.....	1
---	---

*Tomi Haapaniemi*

Reittitiedon louhinta.....	21
----------------------------	----

*Miikka Kivistö*

Vaatimusmäärittelyn rooli yhdistyvien organisaatioiden tietojärjestelmien integraatiossa.....	35
--	----

*Simo Pönni*

Sosiaalisten yhteisöjen etsiminen klusterointimenetelmillä.....	48
---	----

*Ville Simpanen*

Suomen potilastietojärjestelmien käytettävyys.....	61
--	----

*Peter Skogberg*

# Mobiiliavusteisen kielen oppimisen osa-alueet ja tehokkuus

**Tomi Haapaniemi**

## **Tiivistelmä.**

Mobiililaitteiden kehittyminen on tuonut uusia tapoja oppimiseen. Mobiiliavusteinen kielen oppiminen (MALL) on noussut suosituksi ja helpoksi tavaksi oppia vieraita kieliä paikasta ja ajasta riippumatta. Tässä tutkielmassa käydään läpi mobiiliavusteisen kielen oppimisen eri osa-alueita sekä mobiililaitteen avulla tapahtuvan kielen oppimisen tehokkuutta.

**Avainsanat ja -sanonnat:** tietokoneavusteinen kielen oppiminen, mobiiliavusteinen kielen oppiminen, m-learning, mobiilioppiminen

## **1. Johdanto**

Alati kansainvälistyvässä maailmassa on tärkeää pysyä kehityksen mukana ja pyrkiä kouluttautumaan kansainvälisesti aktiiviseksi työntekijäksi ja ihmiseksi. Teknologian ja tietotekniikan kehittyminen on avannut uudenlaiset valtioiden rajat ylittävät markkinat. Tämän seurauksena vieraiden kielten opiskelu ja kulttuureihin tutustuminen ovat nousseet tärkeäksi kouluttautuessaan osaksi kansainvälistä yhteiskuntaa. Teknologiapainotteisessa maailmassa perinteinen luokkahuoneessa tapahtuva opetus siirtyy luokkahuoneen ulkopuolelle mobiililaitteiden avulla tapahtuvaksi oppimiseksi. Mobiilioppiminen on vaivaton ja helppo tapa käyttää lyhyetkin vapaat hetket hyödyksi. Miangahin ja Nezaratin [2012] mukaan kiireiset ihmiset eivät enää ehdi osallistua luokkahuoneopetukseen, vaan he käyttävät mieluummin mobiililaitteita oppimisen apuna.

Mobiiliteknologiasta on tullut tärkeä osa ihmisten jokapäiväistä elämää ympäri maailmaa. Mobiililaitteet vaikuttavat siihen, kuinka ihmiset viestivät keskenään, työskentelevät ja matkustavat muuttuvassa globaalissa maailmassa. Mobiililaitteet mahdollistavat kaiken aikaisen viestinnän, tietoon käsiksi pääsyn ja tiedonhankinnan riippumatta olinpaikasta. Ennen kaikkea, mobiililaitteet tarjoavat nykyään laajat multimediakokemukset ja apukeinot, jotka muuttavat oppimista kaavamaisesta ei-kaavamaiseksi, staattisesta dynaamiseksi ja henkilökohtaisesta yhteisölliseksi [Lam and Cheung, 2010].

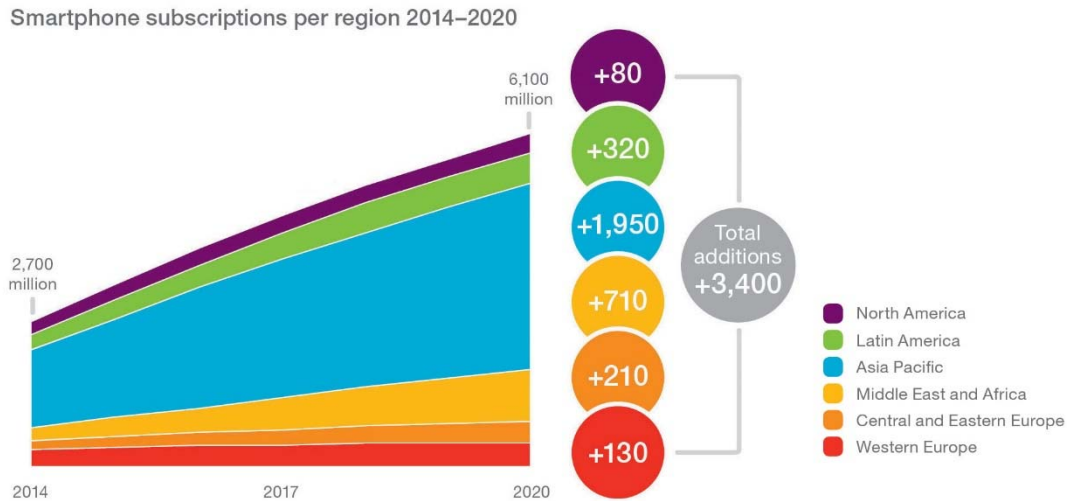
Internet-ympäristö tarjoaa itsenäiseen kielten opiskeluun lukuisia sovelluksia ja materiaaleja, sekä ilmaisia että maksullisia. Mobiilisovellus Duolingo lupaa opettaa oppijan lukemaan, kirjoittamaan, kuuntelemaan ja puhumaan vieraalla

kielellä erittäin tehokkaalla tavalla. Ilmaisen 34 tunnin opiskelun Duolingolla sanotaan vastaavan lukukauden kestäväää ja kallista yliopiston oppimäärää [Duolingo, 2015]. Kielen oppimisalusta Rosetta Stone [2015] puolestaan lupaa, että kielen oppiminen sen avulla on helpompaa kuin ikinä voisi kuvitellakaan. Voiko mobiiliavusteinen kielen oppiminen olla tosiaan niin tehokasta ja helppoa? Tässä tutkielmassa tutkin nimenomaan mobiiliavusteista kielen opiskelua ja sen tehokkuutta kielen oppimisessa. Vaikka mobiililaitteeksi lasketaan usein myös kannettava tietokone, en ota sitä huomioon tässä tutkielmassa, vaan tutkin pienempien ja helposti mukana kannettavien mobiililaitteiden, esimerkiksi tablettien ja älypuhelimien, käyttöä kielen oppimisen välineenä.

Tutkielman tavoitteena on muodostaa kokonaiskuva nykyaikaisesta kielen oppimisesta mobiililaitteiden avulla sekä pohtia mobiilioppimisen tehokkuutta kielen oppimisprosessissa. Tutkielman alussa kerrotaan kehittyneen tekniikan tuomia mahdollisuuksia ja luvussa 3 mobiilioppimisesta ylipäätään. Luvussa 4 käsitellään mobiiliavusteista kielen oppimista ja sen eri osa-alueita. Luvussa 5 perehdytään kahteen eri mobiiliavusteiseen kielen oppimisen sovellukseen ja luvussa 6 mobiiliavusteisen kielen oppimisen tehokkuuteen. Lopuksi pohditaan mobiililaitteen avulla tapahtuvaa kielen oppimisen tehokkuutta.

## **2. Internet-aikakausi**

TNS:n Connected Life -tutkimus [2014] osoittaa, että internet on läsnä suomalaisten arjessa yhä enemmän ja yhä useamman laitteen kautta. Tutkimuksen mukaan suomalaisista 70 prosenttia omistaa älypuhelimien ja 35 prosenttia tabletin. Luku on voimakkaassa kasvussa, sillä 2012 vastaavat luvut olivat 44 prosenttia ja 6 prosenttia [TNS Mobile Life -tutkimus, 2012]. Neljännes kansainväliseen Connected Life 2014 -tutkimukseen osallistujista käyttää tietokonetta, matkapuhelinta tai tablettia joka päivä videoiden katsomiseen internetissä. Varsinkin mobiililaitteet ovat nousseet yhä yleisempään asemaan osaksi jokapäiväistä elämää. Ericssonin mobiliteettiraportti [2014] ennustaa, että 90 prosenttia maailman yli 6-vuotiaista ihmisistä käyttää matkapuhelinta vuonna 2020. Raportin mukaan vuoden 2014 lopussa älypuhelinliittymiä oli maailmassa yhteensä 2,7 miljardia ja vuoteen 2020 mennessä luku tulee tuplaantumaan, kuten kuvassa 1 on esitetty.



Kuva 1: Mobiililiittymisen määrä [Ericsson Mobility Report, 2014]

Nykyään mobiililaitteita, esimerkiksi puhelimia tai tabletteja, käytetään monissa eri käyttötarkoituksissa soittamisesta viestien lähettämiseen, videochattailuun, audioiden kuunteluun, internetissä surffaamiseen ja verkkokaupoissa asioimiseen. Miangah ja Nezarat [2012] painottavat, että viime aikoina niin opettajat kuin opiskelijatkin ovat havainneet mobiililaitteiden sopivan koulutautumiseen ja oppimiseen, sillä niiden avulla oppiminen on niin kaikkialla toteutettavaa kuin mahdollista.

### 3. M-oppiminen

Yksi e-oppimisen alakategorioista on mobiili- eli m-oppiminen. M-oppiminen voidaan määritellä monella eri tavalla. Tarkasta määritelmästä ei ole päästy yksimielisyyteen osittain sen takia, että tieteenala muuttuu ja kasvaa nopeasti. Osa syy on myös se, että mobiilioppiminen toimii parhaiten, kun se on osa jotain muuta alaa. Vaikka oppimisen määritelmä on selkeä, mobiili-sana voi viitata sekä sisällön että oppijan mobiliteettiin eli liikuteltavuuteen. Moura ja Carvalho [2010] korostavat, ettei mobiilioppimisen tavoite ole haastaa tai korvata muita vuorovaikutuksellisia tapoja oppia, vaan toimia täydentävänä tapana, joka voi tukea, rikastuttaa ja tehostaa oppimiskokemusta.

Lyhyesti määriteltynä m-oppiminen on oppimista, joka mahdollistaa poikkitieteellisen sijainnista riippumattoman oppimisen [Lam and Cheung, 2010]. Teknologisesti määriteltynä m-oppiminen on oppimista, jossa käytetään apuna jotain mobiililaitetta. Tällaisia laitteita ovat esimerkiksi kannettavat tietokoneet, MP3- ja muut audiosoitimet, matkapuhelimet ja tabletit [Crompton, 2013]. M-oppiminen on mobiilimpi laajennus e-oppimisesta, koska oppiminen voi tapah-

tua missä tahansa ja mihin aikaan tahansa. Oppijan ei ole pakko olla jossain määritetyssä tilassa tai paikassa, mikä tuo oppijakeskeisen näkökulman oppimiseen [Winters, 2007]. Geddes [2004] lisää määritelmään sanan tieto ja määrittelee m-oppimisen ”minkä tahansa tiedon tai taidon hankkiminen mobiiliteknologiaa hyödyntäen”.

Mobiilioppimisen teknologian kätevyys korostuu luokkahuoneopetuksen ulkopuolella tapahtuvassa oppimisessa. Kukulska-Hulmeen [2009] mukaan opiskelu mobiililaitteella antaa oppijalle mahdollisuuden hyödyntää vapaa-aikaansa oppimisprosessissa ja päättää itse, milloin on aikaa opiskeluun. Lisäksi mobiililaitteella opiskeleminen sopii myös oppimisprosessiin, jossa oppija on liikkeellä paljon, ei ehdi muuten opiskella eikä olla läsnä perinteisessä luokkahuoneopetuksessa.

Sen ohella, että mobiililaitteen voi ottaa helposti mukaan eri ympäristöihin, Klopfer ja muut [2002] liittivät mobiilioppimisen ominaisuuksiin myös sosiaalisen vuorovaikutuksen (mobiililaitteiden avulla oppijat voivat jakaa tietoa ja auttaa toisiaan oppimisessa), asiayhteyden herkkyyden (mobiililaitteen dataa voidaan kerätä sijainnista, ympäristöstä ja ajankohdasta), yhdistettävyyden (voidaan muodostaa jaettua ympäristö verkkoon tai eri laitteiden välille) ja yksilöllisyyden (oppimisalustat voidaan kustomoida yksilöllisesti). Heidän määritelmässään mobiililaitteena oli kämmentietokoneet, joiden suosio ei kuitenkaan ole enää kovin suuri.

El-Husseinin ja Cronjen [2010] mielestä opetuskäytössä mobiilioppiminen on järkevää vain silloin, kun teknologia on täysin mobiilia ja teknologian käyttäjät ovat mobiileja oppiessaan. Mobiliteetilla voidaan siis sekä viitata teknologisiin ominaisuuksiin oppimistilanteessa, mutta toisaalta se voi myös viitata itse oppijaan sekä hänen suorittamaan oppimisprosessiin, käytökseen ja asenteisiin. Mobiililaitteella oppiminen voidaan nähdä oppijakeskeisenä oppimisena, sillä se on joustavaa, helposti saavutettavaa ja henkilökohtaista. Se takaa oppijalle vapauden ja itsenäisyyden oppimiseen. Näiden tekijöiden ansiosta on mahdollista, että oppijan mielenkiinto oppimiseen pysyy yllä ja sen seurauksena oppijan tuoteliaisuus ja tehokkuus paranevat.

#### **4. Mobiiliavusteinen kielen oppiminen**

Tietokoneavusteinen kielen oppiminen, CALL (Computer Assisted Language Learning), on tietokoneavusteisia kielten opettamiseen ja oppimiseen käytettävien sovelluksien tieteenala [Levy, 1997]. Tietokoneavusteinen kielen oppiminen jakautuu alakategorioihin, joista yksi on mobiiliavusteinen kielen oppiminen. Mobiiliavusteinen kielen oppiminen (MALL) tarkoittaa kielen oppimista mobiililaitteen avulla ja se on osa sekä m-oppimista että tietokoneavusteisesta kielen

oppimista. Mobiililaitteita pidetään sopivina varsinkin sosiaalisissa kontakteissa ja yhteisöllisessä oppimisessa, joiden merkitystä pidetään kielen oppimisessa tärkeänä. Kukulska-Hulme ja Shield [2008] erottelevat mobiiliavusteisen kielen oppimisen tietokoneavusteisesta kielen oppimisesta sillä tavalla, että MALL:ssa käytetään omaa henkilökohtaista kannettavaa laitetta, joka mahdollistaa uusia tapoja oppimiseen korostaen jatkuvuutta, käytettävyyden spontaanisuutta ja erilaisten käyttötapojen vuorovaikutusta.

Miangah ja Nezarat [2012] jakavat mobiiliavusteisen kielen oppimisen yleisimpiin kielen oppimisen osa-alueisiin: sanasto, kielioppi, luetun ymmärtäminen, kuullun ymmärtäminen ja fonetiikka eli ääntäminen.

#### **4.1. Sanaston oppiminen**

Yksi osa-alue kielen oppimisessa mobiililaitteella on sanaston opiskelu. Sanaston opiskelu on yksi keskeisimmistä kielen oppimista alueista, koska sanasto luo pohjan lauseiden muodostamiseen ja suulliseen viestimiseen. Song ja Fox [2008] painottavat, että huono sanaston osaaminen voi johtaa väärin päätelmiin tai sisällön väärinymmärryksiin vieraan kielen tekstejä lukiessa. Sen sijaan heidän mukaansa sanakirjojen käytöllä näyttää olevan positiivinen vaikutus opiskelijoiden kielen oppimisessa.

Sanaston oppiminen voidaan jakaa kahteen luokkaan: tarkoitukselliseen - ja tahattomaan oppimiseen. Tarkoituksellinen sanaston oppiminen tarkoittaa oppimista, jossa pyritään pääasiassa parantamaan sanaston oppimista nimenomaan tarkoituksella. Silloin, kun oppiminen on osa lukemista, jossa oppija arvaa tai pääättelee tuntemattomien sanojen merkitystä keskittymättä yksinomaan sanaston oppimiseen, on kyseessä tahaton sanaston oppiminen. Tahaton oppiminen antaa mahdollisuuden päätellä tuntemattoman sanan tarkoituksen asiayhteyden perusteella, jonka seurauksena se on oppijalähtoisempää kuin tarkoituksellinen sanaston oppiminen. Song ja Fox [2008] kuitenkin varoittavat, että tahaton oppiminen voi johtaa väärin päätelmiin, joten oppijalla on jatkuvasti oltava mahdollisuus päästä käyttämään sanakirjaa sanojen tarkistamiseen.

Mobiililaitteisiin on tarjolla lukuisia mobiilisanakirjoja. Niitä voi ladata mobiililaitteisiin tai käyttää mobiililaitteen kautta. Mobiilisanakirjojen käyttö voi olla aikaa säästävää ja niiden avulla voi oppia sanastoa sekä tarkistaa verbien taivutuksia. Rahimi ja Miri [2014] tuovat esiin erilaisia mobiilisanakirjojen käyttötapoihin liittyviä mahdollisuuksia, jotka voivat olla syy onnistuneempaan oppimiseen. Tällaisia ovat muun muassa visuaaliseen mediaan (kuvat), audiomediaan (ääntäminen), multimediaan ja hakuominaisuuksiin liittyvät käyttömahdollisuudet.



## 4.2. Kieliopin opiskelu

Kieliopin osaaminen kielen oppimisessa on tärkeää sekä luetun että kuullun ymmärtämisessä. Vaikka puhekielessä kielioppisäännöt eivät usein pädekään, on tärkeää osata kieltä oikeaoppisesti virallisia tilanteita varten. Kielioppisääntöjä voi oppia mobiililaitteeseen asennetun erillisen sovelluksen avulla, joissa oppijan tehtävä voi olla valita monesta vaihtoehdosta oikea, vastata oikea- tai vääräkysymyksiin tai täyttää tyhjät kohdat itse [Rahimi and Miri, 2014]. Etuna luokahuoneopetukseen on se, että tehtäviä mobiililaitteella tehdessä oppija voi saada välittömän palautteen oikeasta tai väärästä vastauksesta, toisin kuin esimerkiksi perinteisen opetuksen kotitehtävissä.

## 4.3. Luetun ymmärtäminen

Luetun tekstin ymmärtäminen vieraalla kielellä on askel eteenpäin sanaston oppimisesta, kun yksittäisistä sanoista muodostetaan kokonaisia lauseita ja virkkeitä. Sanaston osaamisen lisäksi lukijan täytyy hallita kielioppisääntöjä pystyäkseen ymmärtämään kirjoitetun tekstin sisältöä. Luetun ymmärtäminen voi olla hankalaa, mikäli sanasto ei ole tarpeeksi hyvin hallussa tai ei tiedä kielioppisääntöjä. Kirjoitetun tekstin ymmärtämistä voi parantaa esimerkiksi opettelemalla lisää sanastoa tai totuttautumalla ymmärtämään kieltä lukemalla vieraskielisiä artikkeleita, kirjoja tai vaikka blogeja. Varsinkin nykyään vieraskielisiin teksteihin on helppo päästä käsiksi mobiililaitteella, ja motivoivia omiin mielenkiinnon kohteisiin liittyviä tekstejä on helppo löytää internetistä.

Miangahin ja Nezaratin [2012] mukaan mobiililaitteen avulla luetun ymmärtämisen harjoittelussa voi käyttää erillistä laitteeseen ladattua sovellusta tai tekstiviestipalvelua. Molemmissa tapauksissa oppija saa palautetta omasta luetun ymmärtämisestään. Lisäksi, sellaiset sovellukset, joissa on luetun ymmärtämisen apuna tekstin lausumisäänite mukana, kehittävät samaan aikaan sekä luetun ymmärtämistä että kuullun ymmärtämistä [Miangah ja Nezarat, 2012].

Chen ja muut [2006] toteavat, että luetun ymmärtämisen harjoitukset edistävät oppijan sanastotaitoja, jotka puolestaan nostavat oppijan luetun ymmärtämisen taitoja. He suunnittelivat tutkimustaan varten sovelluksen, jossa käyttäjät lukevat omien tasojensa mukaisia englanninkielisiä uutisia saaden samalla artikkeleihin liittyvää sanastoa oppimisen tueksi. Useat kielen oppimissovellukset painottuvat tekstinkäännöstehtäviin, jossa kehittyy samalla luetun ymmärtäminen ja sanaston osaaminen.

#### 4.4. Kuullun ymmärtäminen

Vieraan kielen kuullun ymmärtäminen on kielen oppimisen kannalta tärkeä taito, jota täytyy kehittää muiden kielen oppimisen osa-alueiden mukana. Vieraalla kielellä viestissä on tärkeä ymmärtää mitä muut ihmiset puhuvat, jotta voi itse osallistua keskusteluun. Kuullun ymmärtäminen voi myös toimia pohjana oikealla lausumisella sekä kielitaidon karttumiselle. Uuden vieraan kielen opiskelussa juuri kuullun ymmärtäminen voi olla oppimisen ensimmäinen vaihe. Varsinkaan nykyajan internet-aikakaudella vieraiden kielten kuulemiselta ei käytännössä voi välttää. Esimerkiksi television ja radion kautta voi jatkuvasti kuulla vieraita kieliä, myös internetistä voi hyvinkin nopeasti ja helposti löytää hyödyllisiä materiaaleja kuullun ymmärtämisen kehittämiseen.

Kuullun ymmärtämisen tekniikat jaetaan kolmeen eri tapaan: alhaalta ylös suuntautuvaan (bottom-up), ylhäältä alas suuntautuvaan (top-down) ja vuorovaikutteiseen malliin. Bottom-up -tekniikassa kuuntelija keskittyy yksittäisiin sanoihin sekä ilmauksiin, ja muodostaa lauseiden merkityksen yhdistämällä yksittäisiä sanoja ja ilmauksia kokonaisuudeksi. Sen sijaan top-down -tekniikkaa käyttävät kuuntelijat yrittävät ymmärtää puheesta ytimen ja tärkeimmät kohdat. Kolmas tekniikka on sekoitus kahta aiempaa tekniikkaa. Se pitää sisällään yksilöllisen, kulttuurisen, sosiaalisen, kontekstuaalisen, tunnepitoisen, strategisen ja kriittisen ulottuvuuden. [Kim, 2013].

Kimin [2013] mukaan taitavaksi kuuntelijaksi kehittyminen vaatii autenttisen ja merkityksellisen harjoitusmateriaalin kuuntelemista. Siitä johtuen sellaisissa maissa, joissa ei voi altistua vieraalle kielelle luokkahuoneen ulkopuolisessa maailmassa, mobiilisovellukset kielen opiskelussa voivat olla hyvinkin käteviä oppimistyökaluja. Suosittuja tapoja kuullun ymmärtämisen kehittämisessä ovat varsinkin erilaiset podcastit. Podcasteja eli esimerkiksi audio- tai videosarjojen jaksoja voi ladata usein ilmaiseksi suoraan mobiililaitteeseen, josta niitä voi kuunnella tai katsoa. Podcastit voivat tarjota oppijalle esimerkkejä autenttisesta ja ”oikeasta” puheesta sekä niitä voidaan käyttää myös tekstmateriaalien lisänä tukemassa oppimisprosessia [Kim, 2013]. Podcastien lisäksi muita mobiiliavusteisia tapoja kuullun ymmärtämisen kehittämisessä ovat esimerkiksi radion kuunteleminen, erilaiset sosiaalisen median kanavat, kuten Skype ja videobloggien seuraaminen, ja kuunteluharjoituksia sisältävät sovellukset.

#### 4.5. Fonetikka

Fonetikka eli ääntäminen on oleellinen osa onnistunutta viestintää. Vieraan kielen puhumisessa tämä varsinkin korostuu, sillä vaikka muut kielen osa-alueet olisivat hyvin hallussa, väärä sanojen ääntäminen voi tarkoittaa sitä, ettei tule

ymmärretyksi viestintätilanteessa. Huono lausuminen johtaa heikkoon viestimistaitoon, joka voi olla kriittistä ei-natiivien kielen puhujien keskuudessa [Segaran et al., 2014]. Kuullun ymmärtämisen sovelluksista voi ammentaa tietoa, kuinka sanoja tulisi ääntää ja kuullun ymmärtämistä harjoitellessa kehittyy varmasti myös oppijan fonetiikkataidot. Puheenohjaus ja puheentunnistus ovat viimeisten vuosien aikana parantuneet merkittävästä ja nykyään on olemassa sovelluksia, joiden avulla voi nimenomaan harjoitella omaa vieraan kielen ääntämistä [Duolingo, 2015; Rosetta Stone, 2015].

Miangahin ja Nezaratin [2012] mielestä hyvän m-oppimispalvelun tulisi sisältää mahdollisuus äänentoistoon, jotta oppija voi ladata oikeaoppisia lausumisoheja mobiililaitteeseensa ja sen jälkeen harjoitella niitä itse. Multimedialaitteissa on mahdollista myös nauhoittaa omaa lausumistaan ja sen jälkeen lähettää lausumiset eteenpäin opettajalle, jolta saa palautteen harjoittelusta. Vaihtoehtoisesti voi myös itse kuunnella omaa lausumistaan ja verraten sitä mallilausumiseen, parantaa omia taitojaan. Podcastit ovat hyödyllisiä kuullun ymmärtämisessä, mutta niitä voi käyttää myös lausumisen oppimisessa. Saatavilla on lukuisia podcasteja, joissa käydään kielen lausumista läpi mallilausujan opastuksella.

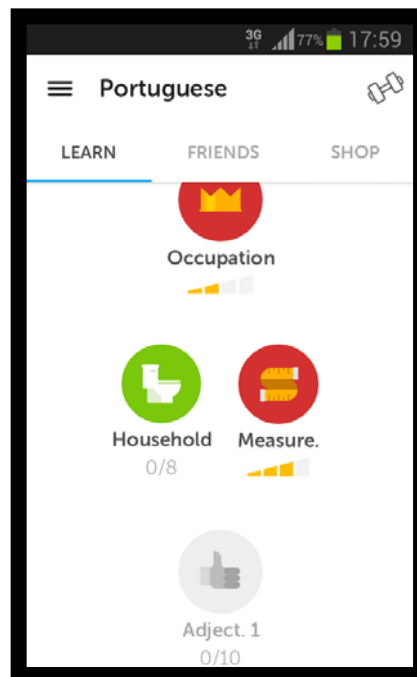
## **5. Mobiiliavusteisen kielen oppimisen sovelluksia**

Valitsin esimerkiksi tutkielmaani varten kaksi mobiiliavusteista kielen oppimisen sovellusta. Toinen sovelluksista on kokonaan ilmainen ja toinen maksullinen. Ilmainen sovellus Duolingo on noussut viime vuosina erittäin suosituksi ja ladatuksi sovellukseksi, minkä takia se on mukana tutkielmassa esimerkkinä. Maksullinen sovellus Rosetta Stone on tietokoneavusteisten kielen oppimisohjelmien pioneeri, joka on saatavilla myös mobiiliversiona.

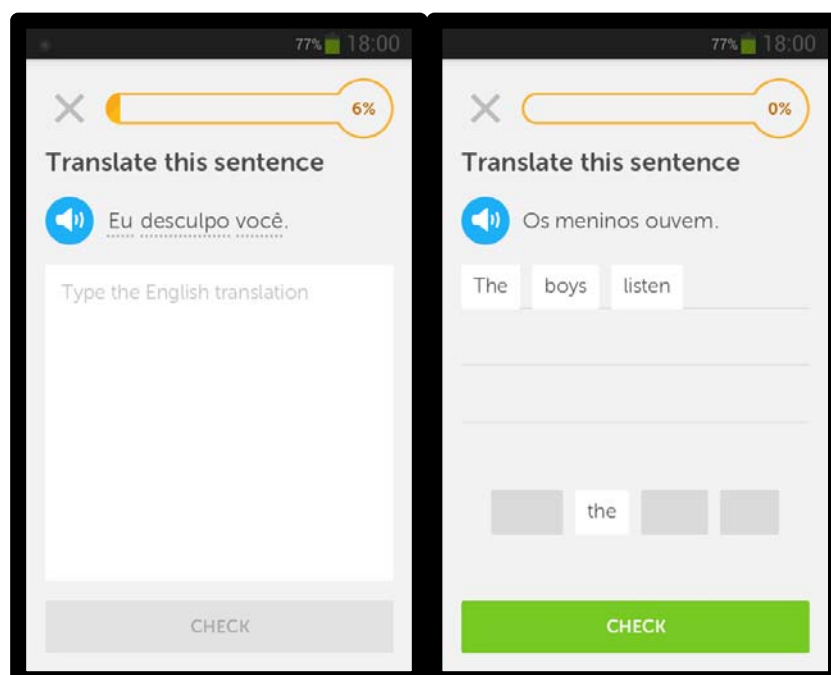
### **5.1. Duolingo**

Yksi tämän hetken suosituimmista kielen oppimispeleistä on kesällä 2012 julkaistu Duolingo. Duolingo, joka on ilmainen online-kielenopiskelupeli, on käytettävissä tietokoneella ja ladattavissa mobiililaitteeseen; Androidille, iOS:lle sekä Windows Phone -alustalle. Duolingolla on yli 80 miljoonaa käyttäjää ja se valittiin iPhonen vuoden sovellukseksi 2013, Googlen Best of the Best -sovellusten joukkoon sekä vuonna 2013 että vuonna 2014 ja TechCrunch -sivuston parhaaksi koulutusalan startupiksi vuonna 2014. Samaan aikaan, kun käyttäjät opiskelevat Duolingon avulla kieliä, heidän kielenopiskeluaan käytetään hyödyksi internetin kääntämisessä. Näin Duolingo on mahdollista pitää ilmaisena sovelluksena käyttäjälle. [Duolingo, 2015]

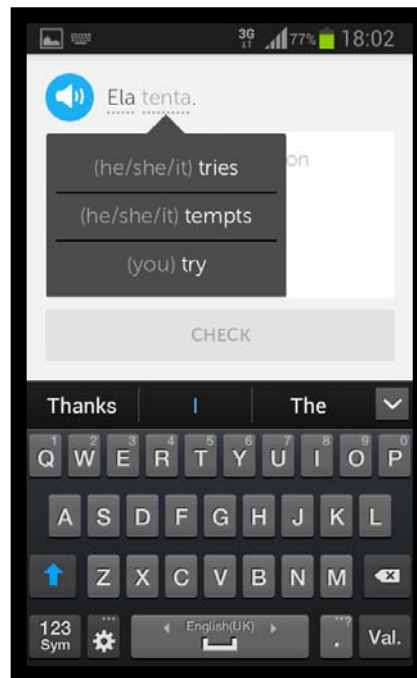
Duolingossa oppija etenee kieltä oppiessaan alaspäin ”taitopuuta”, joka koostuu eri osa-alueisiin tai aiheisiin liittyvistä harjoituksista (kuva 2). Harjoitukset ovat sekoitus erilaisia tehtävätyyppejä ja harjoituksissa voi olla lausumia, kuullun ymmärtämis-, käännös- ja monivalintatehtäviä. Kuten kuva 3 osoittaa, pääpaino Duolingossa on luetun ymmärtämisen harjoituksissa ja useimmiten harjoitukset sisältävät tekstin kääntämistä kielestä toiseen. Kääntäessään lauseita, oppija voi halutessaan saada vihjeitä alleviivatuista sanoista, jonka seurauksena oppija voi kontrolloida omaa oppimisprosessiaan (kuva 4).



Kuva 2: Duolingon taitopuu



Kuva 3: Esimerkkejä Duolington harjoituksista



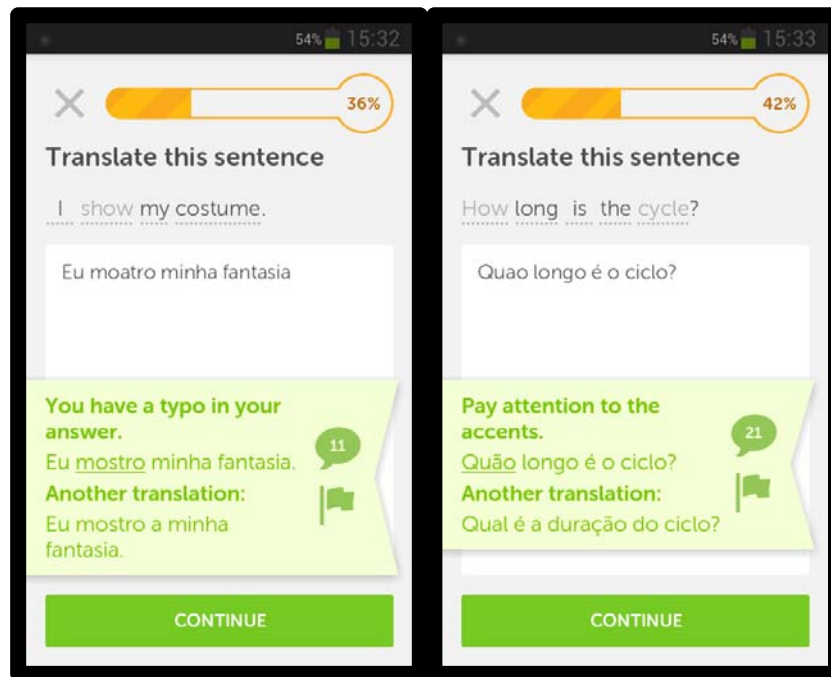
Kuva 4: Duolingo tarjoaa käännösapua käyttäjälle

Kielen oppimisen mielenkiintoisuuden ja sovellusten koukuttavuuden kannalta pelillistäminen on tärkeä osa oppimisprosessin kokonaisuutta ja Duolingoissa se ilmenee monesta eri asiasta. Oppiessaan kieltä oppija saa tekemistään harjoituksista kokempuspisteitä ja etenee tasoittain eteenpäin. Sovellukseen voi asettaa päivittäisen oppimistavoitteen itselleen, mistä sovellus muistuttaa ilmoituksella ja onnittelee tavoitteen ylittyessä. Kun oppija tekee Duolington harjoituksia useana peräkkäisenä päivänä, saa hän palkintomerkkejä, joilla voi "ostaa" muun muassa erilaisia bonusharjoituksia, kuten esimerkiksi opitun kielen flirttailusanaston harjoituksen. Myös sosiaalisuus on tärkeässä asemassa Duolingoissa. Oppijat voivat seurata kavereidensa edistymistä kielen oppimisessa sekä keskustella tehtävistä toisten käyttäjien kanssa. Jokaisella oppijalla on myös maskottipöllö kannustamassa ja motivoimassa oppimiseen. [Duolingo, 2015]

Vesselinov ja Grego [2012] tutkivat kielen oppimisen tehokkuutta kotikielenään englantia puhuvien henkilöiden espanjan kielen oppimisessa Duolington avulla. Tutkimuksessa mitattiin espanjan kielen taitojen parantumista sekä testitulosten että oppimisen kehitykseen käytetyn opiskeluajan suhteen. Tutkimuksen lopputulokset ovat vaikuttavia. Suuri enemmistö tutkimukseen osallistuneista onnistui Duolington avulla parantamaan espanjan kielen taitojaan merkittävästi. Tutkimuksen tulosten perusteella tehdyn johtopäätöksen mukaan espanjan kielen vasta-alkajalta kului keskimäärin 34 tuntia oppia Duolington avulla yliopiston ensimmäisen lukukauden espanjan opinnot. Siis 34 tunnin Duolington

avulla tapahtuva oppiminen vastaisi 11 viikon yliopisto-opetusta. Kielen oppimisen tehokkuuden kannalta tärkeä tekijä tutkimuksessa oli oppijan oma motivaatio kielen oppimiseen. Ne osallistujat, jotka halusivat oppia kieltä matkustelua varten, kehittyivät eniten. Sen sijaan ne osallistujat, jotka opiskelivat enimmäkseen itseään tai koulua varten, oppivat vaatimattomammin. Toinen asia, joka oppimisen tehokkuudesta nousi esiin, oli espanjan kielen taitotaso. Odotevasti aloittelevat kielen opiskelijat kehittyivät eniten, kun taas kokeneemmat espanjan kielen osaajat kehittyivät vain vähän. Myös Garcian [2013] tutkimuksesta ilmenee, että käyttäjien tyytyväisyystaso on sovelluksen käytön alussa korkeimmillaan, josta se alkaa vähitellen hiipua kielen oppimisen edetessä. Näiden faktoiden perusteella voidaan päätellä, että Duolingo saattaisi olla hyvä työkalu pohja- ja perustietojen kartuttamisessa kielen opiskelussa.

Muutenkin Garcia [2013] ottaa hieman kriittisemmän lähestymistavan Duolingoan. Hänen mukaan sovelluksen tarjoamat lauseet voivat välillä olla täysin kömpelöitä sisällöltään, jolloin käyttäjä ei saa esimerkkejä autenttisesta kielestä. Lisäksi, mitä ylemmälle tasolle kielen opiskelussa edetään, sovellus saattaa liputtaa oikean vastauksen aiheettomasti vääräksi. Silti Garcian mukaan Duolingo on ulkoasultaan miellyttävän neutraali hyväksyttäväksi riippumatta käyttäjän kulttuurista tai iästä, sen pelillistäminen toimii hyvin ja se kannustaa yhteistyöhön vertaisten kanssa. Hänen mukaan Duolingon silmiinpistävin ominaisuus on sen kyky antaa aina välittömästi vastauksen syöttämisen jälkeen palaute oppijalle, joka on yleistä muissa sovelluksissa usein vain monivalintatehtävissä. Kuten kuvasta 5 nähdään, Duolingo sen sijaan kertoo käyttäjälle heti, menikö tämän tekemä käännöslause tai -tehtävä oikein vai väärin ja tarjoaa muita oikeita vaihtoehtoja tehtävään tai huomauttaa käyttäjälle mikä vastauksessa meni väärin.

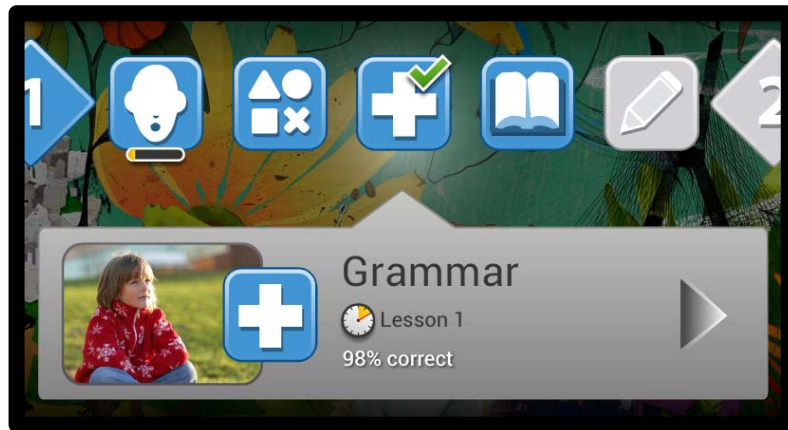


Kuva 5: Duolingon välitön palaute ja vinkit annetusta vastauksesta

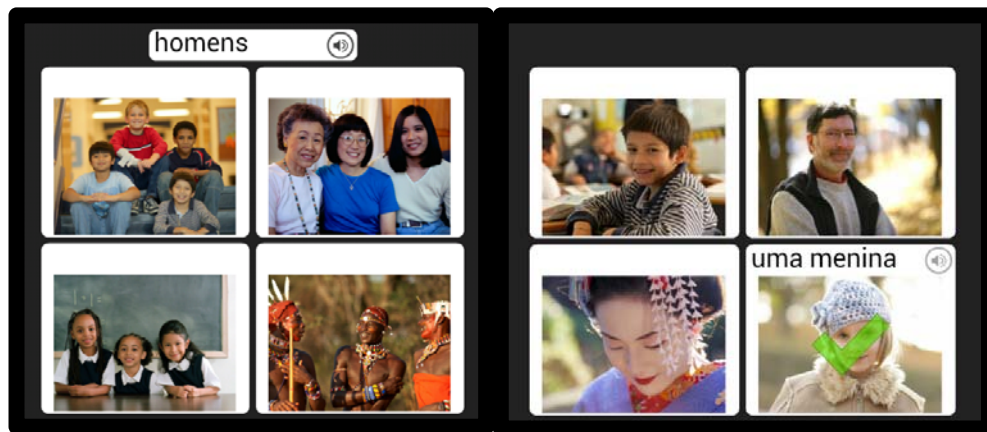
## 5.2. Rosetta Stone

Vuonna 1992 perustettu Rosetta Stone on yksi vuorovaikutteisen kielen oppimisen uranuurtajayrityksistä, ja se on tuhansien koulujen, yritysten ja organisaatioiden sekä miljoonien ihmisten käyttämä. Rosetta Stonen avulla voi opiskella 30 eri kieltä sekä tietokoneeseen asennetulla ohjelmalla, online-versiolla että myös mobiilisovelluksella. Rosetta Stone TOTALe PRO -niminen kielen oppimisen ympäristö jakautuu kolmeen eri osaan. Rosetta World on vuorovaikutteinen online-yhteisö, jossa kielen oppijat ympäri maailmaa voivat harjoitella ja pelata kielenoppimispelejä yhdessä sekä chattailla toistensa kanssa. Rosetta Studio vie kielen oppijat online-keskusteluun kielen ”valmentajan” ja muiden vastaavalla kielen oppimisen tasolla olevien ihmisten kanssa. Live-keskustelu tapahtuu opittavalla kielellä ja takaa näin autenttisen oppimisen vuorovaikutuksen. TOTALe PRO:n pohja on lukuisista kielen oppimisen tasoista koostuva Rosetta Course -oppimisalusta. [Rosetta Stone, 2015]

Rosetta Coursen jokainen taso eli oppitunti pitää sisällään kahdeksan erilaista kielen oppimiseen liittyvää osa-aluetta: lausuminen, sanaston oppiminen, kieliopin oppiminen, luetun ymmärtäminen, kirjoittaminen, luetun ja kirjoittamisen yhteisosio ja puhuminen sekä lisäksi vielä kertausosio (kuva 6). Joissakin tehtävissä oppijan täytyy yhdistää ääni ja/tai teksti oikeaan kuvaan, kuten kuvassa 7, ja joissakin harjoitella ääntämistä tavu kerrallaan malliäänänsä mukaan, kuten kuvassa 8. [Rosetta Stone, 2015]



Kuva 6: Rosetta Stonen eri osa-alueita



Kuva 7: Rosetta Stonen kuvan ja äänen yhdistämistehtäviä



Kuva 8: Sanojen ääntämistä tavuittain Rosetta Stonen malliääntämisen perusteella

Vesselinov [2009] tutki myös Rosetta Stone Course -oppimisalustassa kielen oppimisen tehokkuutta espanjan kieltä opiskelevien henkilöiden oppimista mitaten. Tutkimuksen osallistujat paransivat merkittävästi espanjan kielen taitojaan ja 55 tunnin opiskelu Rosetta Stonea käyttäen vastasi tutkimuksen mukaan yhden yliopiston lukukauden oppimistasoa. Myös suurin osa tutkimukseen osallistuneista paransi suullista espanjan kielen taitoaan. Lähes jokainen osallistuja oli sitä mieltä, että ohjelma oli hyödyllinen, helppokäyttöinen, miellyttävä ja palkitseva. Tutkimuksessa ei ollut kyseessä mobiilisovellus, mutta samankaltaisia tuloksia voidaan olettaa Rosetta Stonen mobiilisovelluksella.



Nielson [2011] päätyy täysin päinvastaisiin tuloksiin tutkimuksessaan, jossa kahtena työpaikan itseoppimisen kohdesovelluksena tutkittiin Rosetta Stonea ja Auralogin TELL ME MOREa. Tutkimukseen osallistuneet olivat vapaaehtoisia aikuisia Yhdysvaltain hallituksen toimistoista. Tutkimuksen suurin esiin tuoma ongelma oli osallistujien suuri kato tutkimuksen aikana, joka Nielsonin mukaan saattoi johtua teknologisista ongelmista tai sovelluksen ohjeistuksen puutteesta. Toisaalta, syynä tutkimuksen suureen keskeyttämisprosenttiin voisi olla myös motivoinnin puute. Mikäli tutkimukseen osallistuvat opiskelivat kieltä töitään eivätkä itseään varten eivätkä myöskään saaneet osallistumisestaan mitään palkkiota, huonot tulokset saattavat olla selitettävissä motivaation puutteella, joka on heikentävä tekijä oppimistuloksille [Berg and Abbas Petersen, 2013; Vesselinov and Grego, 2012].

## 6. Mobiilioppimisen tehokkuudesta

Mobiilioppimisen yksi eduista on yhteisöllisen oppimisen mahdollisuus, joka kannustaa oppimaan. Oppijat voivat jakaa tietoa, taitoja ja asenteitaan vuorovaikutteisesti mobiilioppimisessa. Yhteisöllinen oppiminen auttaa oppijaa tukemaan, motivoimaan ja arvioimaan muita oppijoita saavuttaakseen huomattavan suuren oppimisprosessin. [Miangah and Nesarat, 2012] On todettu, että teknologia voi nostaa oppijan motivaatiota kielen oppimiseen, saattaa luoda vieraan kielen oppimiseen positiivisen asenteen [Rahimi and Hosseini, 2011] ja voi jopa vähentää oppijan ahdistuneisuutta kielen oppitunneilla [Rahimi and Yadollahi, 2011].

Mobiiliteknologian jatkuvan laajentumisen seurauksena koulutus on kehittymässä suuntaan, jossa oppiminen on kokonaisvaltainen ja jatkuva prosessi jokapäiväisessä digitaalisessa toiminnassa. Kuitenkin, esteenä avoimelle ja kaikkialla tapahtuvalla oppimiselle voi olla mobiililaitteiden erilaisista laitteistoista ja ohjelmistoista aiheutuvat ongelmat. Mikäli oppimisalustoja ei ole optimoitu toimimaan jollain tietyllä laitteella, käytettävyys ja käyttäjäkokemus voivat vähentyä. Alhainen käytettävyys saattaa vaarantaa oppimisen tehokkuuden, heikentää oppijan motivaatiota ja jopa aiheuttaa keskeyttämisen koko oppimisprosessille. [Mercurio et al., 2014] Nykyään sovelluksissa ja teknologisissa laitteissa korostuukin yhä enemmän vuorovaikutteisyyden ja käytettävyyden merkitys ennen vallinneen teknologisten elementtien määrän mukaan. Käyttäjät vaativat hyvää käyttöliittymää sekä mukavaa käyttäjäkokemusta. Tällä tavoin myös oppiminen on helpompaa.

### 6.1. Sanaston oppimisen tehokkuus

Kennedy ja Levy [2008] tutkivat Italian kielen sanaston oppimista tekstiviestien avulla. Opiskelijat saivat matkapuhelmiinsa halutessaan luennoilla käydyn sanaston opetusviestejä monta kertaa viikossa, jotta he voisivat kerrata luokkahuoneopiskelua vapaa-aikanaan. Tutkimuksen lopputulokset osoittivat, että tekstiviestit olivat erittäin hyödyllisiä sanaston oppimisessa ja opetuksen tukena.

Rahimin ja Mirin [2014] tutkimuksessa mitattiin mobiililaitesanakirjan vaikutusta kielen opiskelussa. Kolmenkymmenen neljän opiskelijan joukko jaettiin kahteen ryhmään, joista toinen käytti englannin kielen opiskelussa kurssilla apunaan perinteistä sanakirjaa ja toinen mobiililaitteeseen ladattua sanakirjaa. Opettajan ennen tutkimusta ja tutkimuksen jälkeen tekemät testit osoittivat, että mobiililaitetta apunaan käyttäneet opiskelijat suoriutuivat jälkimmäisestä testistä suhteessa aiempaan testiin paremmin ja olivat kehittyneet enemmän kurssin aikana kuin perinteistä sanakirjaa käyttäneet opiskelijat. Tutkimuksen tulokset tukivat aiempia tutkimuksia, joiden mukaan tietokoneavusteisella kielen oppimisen ympäristöllä voi olla positiivinen vaikutus oppijan englannin kielen oppimiseen vieraana kielenä.

### 6.2. Kieliopin opiskelun tehokkuus

Tutkimusten mukaan kieliopin opiskelun mobiililaitteella nähdään positiivisena kielen oppimiskokemuksena. Wang ja Smith [2013] kuitenkin korostavat, että tiettyjen ehtojen tulee täytyä mobiililaitteella kieltä opiskellessa tai muuten kielen oppimisprosessi on rajattua. Nämä heidän tutkimuksestaan ilmenneet ehdot ovat koukuttavien, mutta ei liian pitkien eikä liian vaativien materiaalien tarjoaminen, sopiva määrä opettajan suorittamaa oppimisen seuranta, muiden oppijoiden osallistuminen oppimisprosessiin, kannustinten käyttö oppimisen tukena, yksityisyyden suoja sekä turvallinen mobiilioppimisympäristö.

Nuorille yliopisto-opiskelijoille sopivimpia aiheita kielen oppimiseen ovat erilaisiin kulttuureihin, opiskelijaelämään, vitseihin ja viihdyttäviin tarinoihin liittyvät materiaalit, kun sen sijaan opiskelua merkitsevät kielioppivisat on pidettävä minimimäärässä [Wang and Smith, 2013].

### 6.3. Luetun ymmärtämisen tehokkuus

Lan ja muut [2007] tutkivat kahden erillisen tutkimuksen avulla mobiiliavusteista yhteisöllistä oppimista. Ensimmäisessä tutkimuksessa he tutkivat englantia vieraana kielenä opiskelevien yhteistoimintaa perinteisellä kielen opiskelun tavalla ja siitä esiin nousseita heikkouksia. Toista tutkimusta varten he ottivat tutkimukseen mukaan yhteisöllistä vertaisavusteista mobiilioppimista tukevan

mallin vähentämään ensimmäisessä tutkimuksessa ilmenneitä heikkouksia. Toisen tutkimuksen tulokset osoittivat, että mobiiliavusteinen kielen oppiminen paransi yhteistoimintaa ja vähensi ahdistuneisuutta oppijoiden keskuudessa, nosti heidän lukumotivaatiotansa ja kehitti heidän ääneenlukemistaitoja.

Myös Chen ja muut [2006] päätyivät tutkimuksessaan tulokseen, että mobiiliavusteisuus parantaa oppijan luetun ymmärtämisen taitoja. Heidän mukaan englannin kielisten uutisten lukeminen on hyvin tehokas tapa kehittää oppijan luetun ymmärtämistä. Vaikka tutkimuksessa tutkittiinkin vain englannin kielen opiskelua, on syytä olettaa, että samanlaisia tuloksia voitaisiin saada muidenkin kielten oppimisessa.

Luetun ymmärtämisen oppimisohjelmissa kielen kääntäminen perustuu usein oppimisohjelmaan tallennettuun ennalta käännettyyn tekstiin, jota oppija kääntää ja kone vertaa syötettyä vastausta oikeaan ratkaisuun. Automaattisten käännöskoneiden kielen kääntämisen ominaisuudet ovat vielä niin epävarmalla tasolla, ettei ole suositeltavaa valita mitä tahansa tekstiä ja opetella sen sisältöä käännöskoneen, esimerkiksi Googlen tai Bingin, avulla. Toisaalta joukkoistamista on alettu käyttää apuna kielen kääntämisessä. Internet tarjoaa mahdollisuuden siihen, että joku muu käyttäjä kääntää oppijan haluaman tekstin. Joukkoistamisessa on hyvät puolensa, mutta toisaalta tekstin käännöksen tasosta ei voi olla täysin varma.

#### **6.4. Kuullun ymmärtämisen tehokkuus**

Useat tutkimukset tukevat käsitystä, jonka mukaan mobiililaitteen avulla tapahtuva kuullun ymmärtämisen kehittyminen on tehokasta. Azar ja Nasiri [2014] tutkivat matkapuhelimilla tapahtuvaa kielen oppimista ja sen vaikutuksia, varsinkin kuullun ymmärtämisen tehokkuutta ja tutkimukseen osallistuvien iranilaisten oppijoiden asennetta mobiiliavusteiseen kielen oppimiseen. Tutkimustulokset osoittivat, että mobiililaitetta opiskelun apuna käyttävä ryhmä oli kehittynyt kuullun ymmärtämisessä enemmän testijakson aikana kuin vertailuryhmä ja näin ollen mobiiliavusteisuus toi tehokkuutta kuullun ymmärtämistaitojen kehittymiseen. Lisäksi osallistujat pitivät kielen oppimista puhelimien avulla mielenkiintoisena ja innovatiivisena tapana.

Kim [2013] päätyi vastaavanlaisiin päätelmiin eteläkorealaisten naiscollegeopiskelijoiden kuullun ymmärtämisen kehittymiseen keskittyvässä tutkimuksessaan. Mobiililaiteryhmä paransi testijakson aikana testitulostaan huomattavasti enemmän kuin vertailuryhmä. Voidaan siis todeta, että podcastien käyttö kuullun ymmärtämisen tukena voi olla tehokasta ja parantaa kuullun ymmärtämistaitoja. Lisäksi mobiililaitetta oppimisen apuna käyttävät voivat kontrolloida oppimisprosessiaan paremmin ja haluamallaan tavalla. Tutkimuksen mukaan

podcasteissa esiintyvän autenttisen kielen parissa oppiminen oli vaikeammin ymmärrettävää, mutta myös samaan aikaan hauskeempaa ja motivoivampaa kuin perinteinen luokkahuoneopetus.

### **6.5. Fonetiikan oppimisen tehokkuus**

Ympäröivässä maailmassa vieraan kielen foneettiset taidot kehittyvät väkisin, Suomessa varsinkin englannin kielen taidot, sillä suurin osa ulkomaisista ohjelmista ja elokuvista on englanninkielisiä. Mobiilioppimisessa foneettisia taitoja kehittävät sovellukset kehittyvät kovaa vauhtia niin kuin koko mobiilioppimisen ala. Segaranin ja muiden [2014] tutkimuksesta selviää, kuinka aiemmissa tutkimuksissa multimediatekniikoiden avulla ääntämistä oppineet suoriutuivat testeistä paremmin kuin tietokonetta apunaan käyttäneet. Niinpä he suunnittelivat tutkimustaan varten puhuvan pään sovelluksen lausumisen opettajaksi. Osallistujat olivat tyytyväisiä sovelluksen toimivuuteen, heidän mielestään sovelluksen avulla oli helppo oppia ja he olivat varmoja, että sovellus parantaa heidän lausumistaitojaan.

## **7. Yhteenveto ja pohdintaa**

Kuten useat esitellyt tutkimukset osoittavat, mobiiliavusteinen kielen oppiminen on vahvasti nousussa ja se saavuttaa suurta suosiota perinteisen luokkahuoneopetuksen rinnalla tai jopa sen korvaavana tekijänä. Toisaalta monet tutkielmassa käsitellyt tutkimukset ovat varsin pieniä otoksiltaan, jolloin niiden varaan rakennettuja johtopäätöksiä ei voida pitää täysin merkittävinä. Ala on kokenut suuria muutoksia ja kehittynyt valtavasti harppauksin viimeisten vuosien aikana, joten aivan tuoreiden suurien tutkimusten puuttuminen on selvästi havaittavissa.

Eräs mielenkiintoinen aihe mobiiliavusteiseen kielen oppimiseen on se, voiko se korvata kokonaan perinteisen luokkahuoneopetuksen. Mikäli vielä näin ei ole niin ainakin se tuo selkeästi kielen oppimiseen lisää tehoa muun muassa motivaation kautta. Kenties opettajat voisivat yhdistellä opetuksessaan sekä perinteistä - että mobiiliavusteista kielen oppimista. Mobiililaitteen avulla opiskelu voisi tuoda oppimiseen toivottuja vapauksia tehden oppimisprosessista entistäkin motivoivamman sekä mielenkiintoisemman ja toisaalta opettaja pystyy siten tarkkailemaan oppilaiden oppimisen kehitystä helpommin. Teknologian, kuten esimerkiksi puheentunnistuksen, kehittyessä kielen oppiminen voi olla hyvinkin aito prosessi mobiililaitteella. Sitä paitsi, mobiililaitteen liikuteltavuus ja oppijan vapaus oppia kieltä missä vain ja milloin vain tarjoavat myös kiireisille ihmisille

mahdollisuuden kielen oppimiseen. Ainakin mobiiliavusteisuus näyttäisi toimivan hyvänä lisänä perinteiselle kielen oppimiselle.

Mutta kuinka tosissaan kieltä voi oppia mobiilisovellusten, esimerkiksi Duolington, avulla? Onko kyseessä todellinen kielen oppimisprosessi vai voiko sovelluksista oppia vain matkusteluun kelpaavaa perustaitoa? Toisaalta perustaitojen avulla omaa kielen osaamistaan voi helposti syventää ja tärkeintä perusviestimisessä vieraalla kielellä voikin olla, että ymmärtää ja tulee ymmärretyksi. Duolington tutkimus [Vesselinov and Grego, 2012] osoittaa myös sen, että oppijan oma motivaatio on erittäin suuressa roolissa vieraan kielen opiskelussa. Pakollisessa kouluopetuksessa tämä tarkoittaisi sitä, että opettajien tulisi pitää oppilaat täysin motivoituneina oppijoina, jotta nämä saisivat itsestään eniten irti. Oppiminen on motivoivampaa ja tehokkaampaa, kun se osataan suunnata oppijan psykologisiin tarpeisiin liittyväksi [Berg and Abbas Petersen., 2013]. Näin oppimisprosessista on syytä tehdä mahdollisimman hauska, jotta kaikki voisivat nauttia kielen oppimisesta.

## Viiteluettelo

- [Azar and Nasiri, 2014] Ali Sorayyaei Azar, and Hassan Nasiri, Learners' Attitudes toward the Effectiveness of Mobile Assisted Language Learning (MALL) in L2 Listening Comprehension. In: *Proc. of Social and Behavioral Sciences* **98** (2014), 1836-1843.
- [Berg and Abbas Petersen, 2013] Malin Aas Berg, and Sobah Abbas Petersen, Exploiting Psychological Needs to Increase Motivation for Learning. *Serious Games Development and Applications*, (2013), 260-265.
- [Chen et al., 2006] Chih-Ming Chen, Shih-Hsun Hsu, Yi-Lun Li, and Chi-Jui Peng, Personalized Intelligent M-learning System for Supporting Effective English Learning. In: *Proc. of IEEE International Conference on Systems, Man and Cybernetics, SMC '06*. **6** (2006), IEEE, 4898-4903.
- [Crompton, 2013] Helen Crompton, A Historical Overview of Mobile Learning: Toward Learner-centered Education. In: *Handbook of Mobile Learning* (2013), 3-14.
- [Duolingo, 2015] Duolingo, Available as <https://www.duolingo.com/info>, Checked 01.05.2015.
- [El-Hussein and Cronje, 2010] Mohamed Osman M. El-Hussein, and Johannes C. Cronje, Defining Mobile Learning in the Higher Education Landscape. *Educational Technology & Society* **13**, **3** (2010), 12-21.
- [Ericsson Mobility Report, 2014] <http://www.ericsson.com/ericsson-mobility-report>, <http://www.ericsson.com/res/docs/2015/ericsson-mobility-report-feb-2015-interim.pdf>

- [Garcia, 2013] Ignacio Garcia, Learning a Language for Free While Translating the Web. Does Duolingo Work? *International Journal of English Linguistics* 3, 1 (2013), 19-25.
- [Geddes, 2004] S., Mobile Learning in the 21st Century: Benefit for Learners. *Knowledge Tree e-journal* 30, 3 (2004), 214-228.
- [Kennedy and Levy, 2008] Claire Kennedy, and Mike Levy, L'italiano al Telefonino: Using SMS to Support Beginners' Language Learning. *ReCALL* 20, 3 (2008), 315-330.
- [Kim, 2013] Hea-Suk Kim, Emerging Mobile Apps to Improve English Listening Skills. *Multimedia-Assisted Language Learning* 16, 2 (2013), 11-30.
- [Klopfer et al., 2002] Eric Klopfer, Kurt Squire, and Henry Jenkins, Environmental Detectives: PDAs as a Window into a Virtual Simulated World. In: *Proc. of IEEE International Workshop on Wireless and Mobile Technologies in Education* (2002), 95-98.
- [Kukulska-Hulme, 2009] Agnes Kukulska-Hulme, Will Mobile Learning Change Language Learning? *ReCALL* 21, 2 (2009), 157-165.
- [Kukulska-Hulme and Shield, 2008] Agnes Kukulska-Hulme, and Lesley Shield, An Overview of Mobile Assisted Language Learning: From Content Delivery to Supported Collaboration and Interaction. *ReCALL* 20, 3 (2008), 271-289.
- [Lam et al., 2010] Jeanne Lam, Jane Yau, and Simon K.S. Cheung, A Review of Mobile Learning in the Mobile Age. *Lecture Notes in Computer Science* **6248** (2010), 306-315.
- [Lan et al., 2007] Yu-Ju Lan, Yao-Ting Sung, and Kuo-En Chang, A Mobile-Device-Supported Peer-Assisted Learning System For Collaborative Early EFL Reading. *Language Learning & Technology* 11, 3 (2007), 130-151.
- [Levy, 1997] Michael Levy, *Computer-Assisted Language Learning: Context and Conceptualization*. Oxford: Clarendon Press, 1997.
- [Mercurio et al., 2014] Marco Mercurio, Ilaria Torre and Simone Torsani, Responsive Web and Adaptive Web for Open and Ubiquitous Learning. *Open Learning and Teaching in Educational Communities* **8719** (2014), 452-457.
- [Miangah and Nezarat, 2012] Tayebah Mosavi Miangah, and Amin Nezarat, Mobile-Assisted Language Learning. *International Journal of Distributed and Parallel Systems* 3, 1 (2012), 309-319.
- [Moura and Carvalho, 2010] Adelina Moura, and Ana Amélia Carvalho, Mobile Learning: Using SMS in Educational Contexts. *IFIP Advances in Information and Communication Technology* **324** (2010), 281-291.

- [Nielson, 2011] Katharine B. Nielson, Self-study with Language Learning Software in the Workplace: What happens. *Language Learning & Technology* 15, 3 (2011), 110-129.
- [Rahimi and Miri, 2014] Mehrak Rahimi, and Seyed Shahab Miri, The Impact of Mobile Dictionary Use on Language Learning. *Procedia - Social and Behavioral Sciences* 98 (2014), 1469-1474.
- [Rahimi and Hosseini, 2011] Mehrak Rahimi, and S. Fatemed Hosseini K., The Impact of Computer-Based Activities on Iranian High-School Students' Attitudes Towards Computer-Assisted Language Learning. In: *Proc. Computer Science* 3 (2011), 183-190.
- [Rahimi and Yadollahi, 2011] Mehrak Rahimi, and Samaneh Yadollahi, Success in Learning English as a Foreign Language as a Predictor of Computer Anxiety. In: *Proc. Computer Science* 3 (2011), 175-182.
- [Rosetta Stone, 2015] Rosetta Stone, Available as <https://www.rosettastone.eu/press/fact-sheets>, Checked 11.5.2015.
- [Segaran et al., 2014] Kogilathah Segaran, Ahmad Zamzuri Mohamad Ali, and Tan Wee Hoe, Usability and User Satisfaction of 3D Talking-Head Mobile Assisted Language Learning (MALL) App for Non-Native Speakers. In: *Proc. Social and Behavioral Sciences* 131 (2014), 4-10.
- [Song and Fox, 2008] Yanjie Song, and Robert Fox, Integrating Incidental Vocabulary Learning Using PDAs into Academic Studies: Undergraduate Student Experiences. *Lecture Notes in Computer Science* 5169 (2008), 238-249.
- [TNS Connected Life -tutkimus, 2014]  
<http://www.tns-gallup.fi/uutiset.php?aid=15056&k=14320>, kansainvälinen <http://connectedlife.tnsglobal.com>,  
<http://www.tnsglobal.com/press-release/connected-life-tv-press-release>
- [TNS Mobile Life -tutkimus, 2012]  
<http://www.tns-gallup.fi/uutiset.php?aid=14815&k=14320>
- [Vesselinov, 2009] Roumen Vesselinov, Measuring the Effectiveness of Rosetta Stone. *City University of New York, USA* (2009).
- [Vesselinov and Grego, 2012] Roumen Vesselinov, and John Grego, Duolingo Effectiveness Study. *City University of New York, USA* (2012).
- [Wang and Smith, 2013] Shudong Wang, and Simon Smith, Reading and Grammar Learning Through Mobile Phones. *Language Learning & Technology* 17, 3 (2013), 117-134.
- [Winters, 2007] Niall Winters, What Is Mobile Learning. In: *Big Issues in Mobile Learning* (2007), 7-11.

## Reittitiedon louhinta

**Miikka Kivistö**

### **Tiivistelmä.**

Tarkastelen tässä tutkielmassa paikantimien tallentaman tiedon louhintaa. Keskityn erityisesti louhitun tiedon käyttöön enduroharrastuksessa ja reittien vaihtoon erikoistuneissa palveluissa.

**Avainsanat: GPS, tiedonlouhinta.**

### **1. Johdanto**

Paikantimien käyttö niin suunnistuksen kuin ulkoiluunkin liittyvien aktiviteettien apulaitteina on yleistynyt jo pari vuosikymmentä. Moni kantaa paikanninta mukanaan jo peruspuhelimen varusteena, autoissa ja moottoripyörissä navigaattorit ovat myös vakiovarustusta. Kuntoilussa on omat juuri tuohon tarkoitukseen kehitetyt tuotteet. Yhä lisääntyvä paikannuslaitteiden käyttö ja kehitys reittipalveluiden tarjonnassa ja reittipalveluihin keskittyvissä verkkopalveluissa luokannustavan toimintaympäristön paikannukseen liittyvällä tietojenkäsittelyn alalla.

Paikanninlaitteiden tallentamia reittejä voi jakaa lukuisissa verkkopalveluissa muille käyttäjille. Reitteihin voi tällöin liittää kuvaa ja muuta informaatiota, jota kukin käyttäjä katsoo kiinnostavaksi tai tarpeelliseksi muille jakaa. Tiedon lisääminen pelkkään reittiin on yleensä muiden apulaitteiden varassa ja edellyttää ainakin vielä erillistä ohjelmaa, jolla reitin voi jakaa muille käyttäjille verkossa. Muun kuin pelkän reitin jakaminen edellyttää manuaalista kuvien ja tekstin syöttämistä reittitiedon lomaan ja voi olla melko hidasta ja vaivalloista.

Tiedonlouhinnan prosesseilla paikantimen tallentamasta reitistä voidaan loushia lisää käyttäjien kiinnostuksen kohteita ja reitin luonteesta lisäinformaatiota. Tässä tutkielmassa hahmotellaan tietämyksen muodostamisen eri prosessien mahdollisuuksia reittitiedon analysoinnissa. Tavoitteena on kartoittaa niitä mahdollisuuksia, joilla reittipalveluiden reittitietojen esittämistä voitaisiin automatisoida ja täten helpottaa reittien havainnollistamista. Erityishuomio on enduroharrastuksella ja reittienvaihtoon erikoistuneilla palveluilla verkossa. Käyn läpi lupaavimpia vaihtoehtoja reittitiedon analysoinnissa tiedon louhinnan prosessien avulla. Tukena esitän tutkielmaa varten moottoripyörällä kerätyn reittitie-



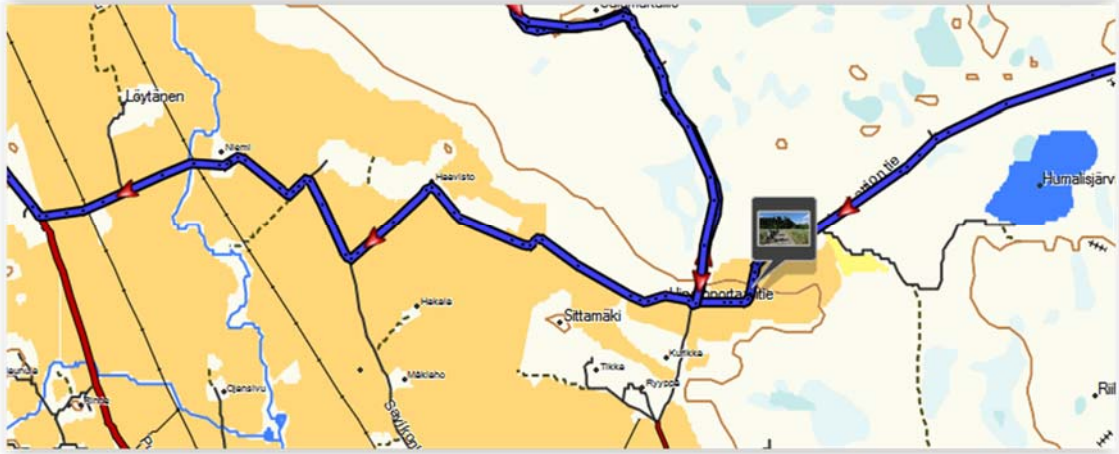
don ja siitä tiedonlouhinnan prosessien avulla saatuja tuloksia ja mahdollisuuksia reittitiedon louhinnassa. Nopeuden ohella, nopeuden vaihtelut, matkan ja ajan suhteen sekä kurssimuutoksien vastaavat vaihtelut osoittivat lupaavia mahdollisuuksia reittitiedon analysoimisessa ulkoiluun liittyvien aktiviteettien kannalta.

Luvussa 2 käydään läpi reittien vaihdon peruskäsitteitä sekä esitellään reittien vaihtoon keskittyvien verkkopalveluiden tarjoamia palveluja ja mahdollisuuksia kehittää näitä palveluja tiedon louhinnan prosessien avulla. Luvussa 3 luonnehditaan tutkielman tueksi kerättyä reittitietoaineistoa, ja esitellään yleisellä tasolla tiedonlouhinnan menetelmiä reittitiedon tutkinnassa. Luvussa 4 esitellään reittitykseen keskittyvien tutkimuksien tuloksia sekä tutkielmaa varten kerätyn aineiston kannalta parhaita tiedonlouhintamenetelmiä ja tutkielman aineistosta louhittuja havaintoja.

## **2. Peruskäsitteitä**

Reittien vaihtoon erikoistuneissa verkkopalveluissa käyttäjät voivat jakaa GPS-laitteella tallentamia reittejä muiden saataville tai voivat etsiä muiden käyttäjien tallentamia ja verkkopalveluun siirrettyjä reittejä omaan käyttöönsä. Verkkopalvelujen reittien vaihdolla voidaan löytää uusia kiinnostavia paikkoja, tai vaikka kokonainen uusi kiinnostava ja kokemuksen arvoinen reitti. Reittejä etsittäessä palveluihin annetaan yleensä syötteenä sijainti ja käyttäjää kiinnostavat asiat, jolloin palvelu voi näyttää kartalla lähistöllä sijaitsevat kiinnostusta vastaavat kohteet, järjestäen ne osuvimmasta alkaen. Joissain sovelluksissa käyttäjä voi antaa myös palautetta tuloksesta antamalla hakutuloksen kohteille arvosanan.

Verkkopalveluiden reitit on usein jaoteltu erilaisiin kategorioihin, kuten polkupyöräily, moottoripyöräily tai vaellus. Basecamp-ohjelmistossa, johon liittyy verkkopalvelun mahdollisuus reittien vaihdossa, kategoriointi erilaisiin aktiviteetteihin tehdään käyttäjän toimesta manuaalisesti ladattaessa reitti palveluun. Reitteihin voi lisätä vihjeitä, kuvia, videolinkin Youtube-videopalveluun, vaikeustason, maamerkkejä, sanallisia kuvauksia eri paikoista ja muuta informaatiota reitin luonteen ymmärtämisen helpottamiseksi. Karttanäkymä Basecamp-ohjelman reitistä on esitetty kuvassa 1. Reittiin on liitetty tällä osuudella yksi kuva reitin varrelta, jonka voi klikata suuremmaksi katselua varten.

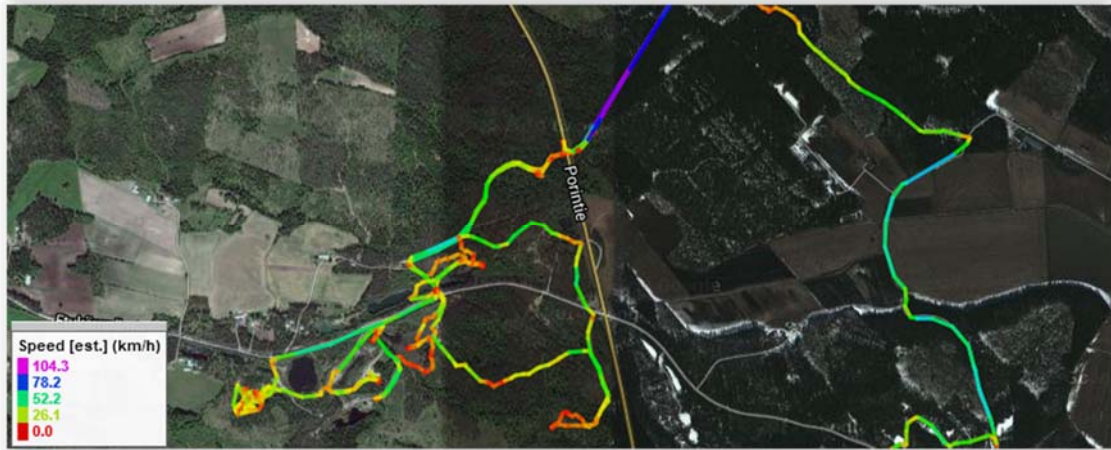


Kuva1. Basecamp-ohjelman karttanäkymä.

Tiedon lisääminen reittiin on kuitenkin melko työlästä ja myös reitin luonteen ymmärtäminen, tiedon hahmottaminen reitistä, vaatii kaikkien siihen lisättyjen kuvien ja vihjeiden läpikäymistä. Sinänsä hienoja matkakertomuksia palvelulla voi toki tehdä.

Tiedonlouhinnan prosessien avulla voitaisiin reittitiedon havainnollistamista verkkopalveluissa automatisoida. Esimerkiksi GPSVisualizer-ohjelma näyttää tallennetun reitin Googlen kartalla, väritettynä nopeuden mukaan. Ohjelmalla vedostettu reitti on esitetty kuvassa 2. GPSVisualizer tarjoaa jo huomattavan paljon informatiivisemman kuvan reitistä vaikkapa enduroharrastuksesta kuntoilu-mielessä kiinnostuneelle käyttäjälle kuin pelkän yksivärisen reitin esittäminen kartalla. Muuta havainnollistamisen mahdollisuuksia ohjelma ei tarjoa, ei myöskään verkkopalvelua reittien vaihtoon. Reittien havainnollistaminen GPSVisualizerin tavalla vielä pidemmälle vietyä ja toiminnon yhdistäminen ohjelmistoihin, jotka mahdollistavat reittien vaihdon, toisi monelle käyttäjälle helpotusta reittien etsimiseen palvelusta kuten sen tallentamiseenkin.

Pidemmälle menevässä havainnollistamisessa reitti voitaisiin esimerkiksi kategorioida automaattisesti. Myös eri kategorioiden sisällä reitin vaikeusastetta voitaisiin esittää esimerkiksi värittämällä reitin eri osat eri väreillä vaikkapa vaikeusasteen mukaan tai vedostamalla reitistä diagrammi, joka havainnollistaa joi-tain sen ominaisuutta. Tässä tutkielmassa hahmotellaan tietämyksen muodostamisen eri prosessien mahdollisuuksia reittitiedon analysoinnissa. Tavoitteena on kartoittaa niitä mahdollisuuksia, joilla reittipalveluiden reittitietojen esittämistä voitaisiin automatisoida ja täten helpottaa reittien havainnollistamista.



Kuva2. GPSVisualizer-ohjelmalla vedostettu reitti.

### 3. Reitit ja niiden tutkiminen

GPS-laitteen tallentaman reittitiedon sisältämää informaatiota selvennetään tämän luvun kohdassa 3.1. Tutkielman tueksi kerättyä reittitieto aineistoa luonnehditaan kohdassa 3.2, jossa myös esitetään muuta reittitietoon liittyvää ja tutkimisen arvoista informaatiota. Kohdassa 3.3 esitetään tiedonlouhinnan menetelmiä reittitiedon tutkinnassa.

#### 3.1 GPS-piste, paikkatieto, sijaintitieto ja reitti

Reitit koostuvat yksittäisistä GPS-laitteen tallentamista koordinaattipisteistä, eli GPS-pisteistä, joihin on liitetty aikaleima ja jotka aikaleimojen määräämässä ajallisessa järjestyksessä muodostavat reitin. Muuta laitteen keräämää tietoa on esimerkiksi korkeustieto. Laite tai ohjelmisto pystyy myös laskemaan kahden peräkkäisen koordinaattipisteen välisen etäisyyden koordinaattien perusteella sekä nopeuden pisteestä A pisteeseen B etäisyyden ja aikaleimojen avulla. Nopeuden tavoin myös kurssi eli suuntima on johdettua tietoa ainakin edistyneemmissä laitteissa. Kuvassa 3 on Basecamp-ohjelmiston yhteenveto ja osa reittipisteistä Garminin GPS-laitteen keräämästä sijaintitietodatasta, jossa yksittäinen rivi edustaa yhtä paikkatietoa sisältäen sijaintitiedon [Kallioniemi, 2012: luku 2]. Rivit yhdessä muodostavat käyttäjän kulkeman reitin, jota on havainnollistettu vektorimuodossa kuvassa 1, yhdistettynä karttaan.

Yhteenveto		Aika		Nopeus		Korkeus merenpinnasta	
Pisteet:	27	Kulunut aika:	0:01:45	Keskimäärin:	36.8 km/h	Minimi:	41 m
Matka:	1.1 km	Aika liikkeellä:	0:01:45	Keskimäärin liikkeellä:	36.8 km/h	Nousu:	11 m
Ala:	84657 m <sup>2</sup>	Aika pysähdyksissä:	0:00:00	Minimi:	15 km/h	Maksimi:	54 m
				Maksimi:	56 km/h	Lasku:	22 m
						Kaltevuus:	-1.1 %

Hakemisto	Korkeus merenpinnasta	Etapin etäisyys	Etappiaika	Etapin nopeus	Etapin kurssi	Aika	Sijainti
1	53 m	11 m	0:00:02	20 km/h	333.5° tosi	17.10.2014 13:24:33	N61° 17.058' E21° 47.759'
2	54 m	21 m	0:00:05	15 km/h	345.9° tosi	17.10.2014 13:24:35	N61° 17.063' E21° 47.753'
3	54 m	35 m	0:00:06	21 km/h	344.4° tosi	17.10.2014 13:24:40	N61° 17.074' E21° 47.748'
4	54 m	11 m	0:00:01	39 km/h	339.0° tosi	17.10.2014 13:24:46	N61° 17.092' E21° 47.737'
5	54 m	94 m	0:00:09	38 km/h	345.6° tosi	17.10.2014 13:24:47	N61° 17.098' E21° 47.733'
6	51 m	21 m	0:00:02	37 km/h	343.3° tosi	17.10.2014 13:24:56	N61° 17.147' E21° 47.706'
7	49 m	90 m	0:00:09	36 km/h	328.5° tosi	17.10.2014 13:24:58	N61° 17.157' E21° 47.700'
8	44 m	29 m	0:00:03	35 km/h	318.5° tosi	17.10.2014 13:25:07	N61° 17.199' E21° 47.647'
9	43 m	23 m	0:00:03	27 km/h	332.3° tosi	17.10.2014 13:25:10	N61° 17.211' E21° 47.625'
10	44 m	57 m	0:00:07	29 km/h	1.6° tosi	17.10.2014 13:25:13	N61° 17.222' E21° 47.613'
11	47 m	26 m	0:00:03	31 km/h	358.9° tosi	17.10.2014 13:25:20	N61° 17.252' E21° 47.615'
12	49 m	22 m	0:00:03	26 km/h	341.9° tosi	17.10.2014 13:25:23	N61° 17.266' E21° 47.615'
13	49 m	37 m	0:00:05	27 km/h	286.1° tosi	17.10.2014 13:25:26	N61° 17.277' E21° 47.607'

Kuva3. Basecamp-ohjelman reittipisteet.

### 3.2 Datan lähteet reitityksen tutkimuksessa

GPS-laitteiden tallentama reittijälki koordinaatteineen, aikaleimoineen ja niistä johdettuine suureineen ovat luonnollisesti tärkein data analysoitaessa reittitietoa. Myös kommentteista, joita käyttäjä on liittänyt reittipisteeseen, voidaan saada lisäinformaatiota. Kommentoituja ja merkattuja pisteitä kutsutaan termillä Point of Interest (POI). Tiettyä aluetta, joka sisältää useita POI-merkintöjä, kutsutaan termillä Region of Interest (ROI). Esimerkiksi geokätkentä sovellukset käyttävät hyväkseen POI-merkintöjä, joita voi ladata useista verkkopalveluista GPS-laitteeseen. Myös nykyaikaiset kamerat pystyvät lisäämään kuvaan sijaintitiedon. GPS-laitteella kuva voidaan lisätä osaksi reitin POI-pistettä ja edelleen jakaa reitti verkossa muille käyttäjille.

Keskityin tässä tutkielmassa enduropyörän maastossa ajon analysoimisen mahdollisuuksien tarkasteluun. Enduropyörällä ja Garminin GPS-laitteella kolmesta erilaisesta maastosta kerätystä aineistosta yritän löytää mahdollisuuksia analysoida reittijälkeä tiedonlouhinnan keinoin. Aineistona on valittu kolme eri segmenttiä, joita kutsun polku-, polkutie- ja tiesegmenteiksi. Kutakin segmenttiä edustavaan aineistoon on valittu alle kahden kilometrin pituinen reitti. Polkusegmentti sisältää 164 sijaintitietoa, polkutiesegmentti 25 ja tiesegmentti 10. GPS-laite rekisteröi paikkatiedon tietyin väliajoin, jolloin paikkatieto kahden kilometrin matkalla hyvin erilaisilla maastonpohjilla rekisteröityy ajallisesti tasaisesti, mutta hyvin erilaisin välimatkoin.

### 3.3 Reittitiedon tiedonlouhimis-menetelmiä

Reittitiedon analysointiin on käytetty monenlaisia tiedonlouhimismenetelmiä. Luokittelumalleilla (classification model) voidaan useista GPS-laitteiden tallentamista reiteistä mallintaa luokittelu malleja, joilla voidaan ennustaa ja kategorisoida vastaavia reittejä. Luokittelumalleja voidaan käyttää reittien segmentoinnissa. Klusterianalyysillä voidaan etsiä esimerkiksi toistuvia reittejä. Poikkeavien tapausten analyysillä (outlier-analysis) epätyypillisiä reitin osia voidaan identifioida. Menetelmällä voidaan löytää datapisteitä tai joukkoja, jotka poikkeavat datan yleisistä ominaisuuksista riittävästi.

Sekvenssianalyysillä (sequential pattern mining) erinäisiä ominaisuuksia voidaan tunnistaa reiteistä ja erotella reitin eri osien tyypillisiä ominaisuuksia toisistaan.

## 4. Reittitiedon tutkinta; mitä voidaan tutkia

### 4.1 Toistuvat reitit ja paikat

Toistuvan reitin louhimisen ongelmana on, ettei seurattava kohde toista reittiään juuri samaan aikaan ja juuri samoilla koordinaateilla. Reitit (periodit) voivat myös lomittua toistensa kanssa, reitti voi toistua päivittäin ja toinen viikoittain. Toistuvissa reiteissä reittien aikaväli on myös tuntematon. Eläinten käyttäytymiseen keskittyvässä metodissa [Li et al., 2011] aikavälin tunnistamisessa on käytetty menetelmää, jossa reittitieto muutetaan binäärikoodiksi, jossa arvo 1 annetaan pisteelle, joka kytkeytyy tiettyyn paikkaan esiintyen toistuvasti datassa, ja 0 muille pisteille. Toistuvat paikat voidaan tunnistaa jakamalla tila koordinaattipisteiden matriisiksi halutulla resoluutiolla ja etsien tuosta matriisista tiheytyksiä, siis paikkoja, joissa usein käydään ja viivytään. Toistuvien reittien tutkiminen on kiinnostavaa myös luonnossa liikkujan kannalta. Menetelmällä voidaan etsiä esimerkiksi vaeltajien ja muiden luonnossa liikkujien reittivalintoja. Sama koskee myös vaikkapa liikennettä. Missä pysähtyytään, miten pitkiä päivämatkoja tehdään ja mitä kautta tavoite saavutetaan. Reittitiedon binäärikoodiksi muuntamista voidaan käyttää myös havainnollistamaan esimerkiksi kiihdytysten ja jarrutusten osuutta reitin erilaisilla segmenteilla. Menetelmää on kuvattu alakohdassa 4.3.2.

## 4.2 Viipymispiste ja viipymisalue

Viipymispiste on maantieteellinen alue, jonka sisällä käyttäjä viipyy tietyn määritellyn ajan. Viipymispiste lasketaan reittipisteiden välisen matkan ja ajan perusteella, joissa matka ja aika pysyvät määritellyjä vakioita pienempinä. Viipymispisteet voivat viitata samaan kiinnostavaan alueeseen käyttäjien reittitiedoissa. Jos samalla alueella sijaitsee usea viipymispiste, voidaan tällaista aluetta kutsua viipymisalueeksi. Käyttäjien tarkkaillessa esimerkiksi jotain maamerkkiä voivat he tarkkailla sitä eri paikoista. Tarkkailupaikkojen joukko edustaa siis viipymisaluetta ja eri tarkkailupaikat viipymispisteitä tuon alueen sisällä.

Klusterioimalla voidaan viipymispisteitä jaotella kuuluvaksi tiettyyn viipymisalueeseen. Zheng ja muut [2010] käyttävät kehittämänsä ruudukkopohjaista algoritmia etsiessään viipymisalueita reittitiedoista. Yhdistettäessä saadut alueet tietoon alueen aktiviteeteista ja käyttäjän kiinnostuksen kohteista voidaan tuloksia käyttää ehdotettaessa toiselle, samalla alueella olevalle, ja samoista asioista kiinnostuneelle käyttäjälle sopivia kohteita tietyllä alueella. Menetelmällä voidaan siis automatisoida POI-pisteiden ja ROI-alueiden muodostusta, sekä yhdistää alueet tiettyihin aktiviteetteihin ilman käyttäjien manuaalisesti reittitietoon lisäämää informaatiota juuri tuon alueen aktiviteeteista. Käyttäjien kiinnostuksen kohteet ja alueella sijaitsevat niitä vastaavat aktiviteetit on etsitty tuossa tutkimuksessa erillisestä datasta.

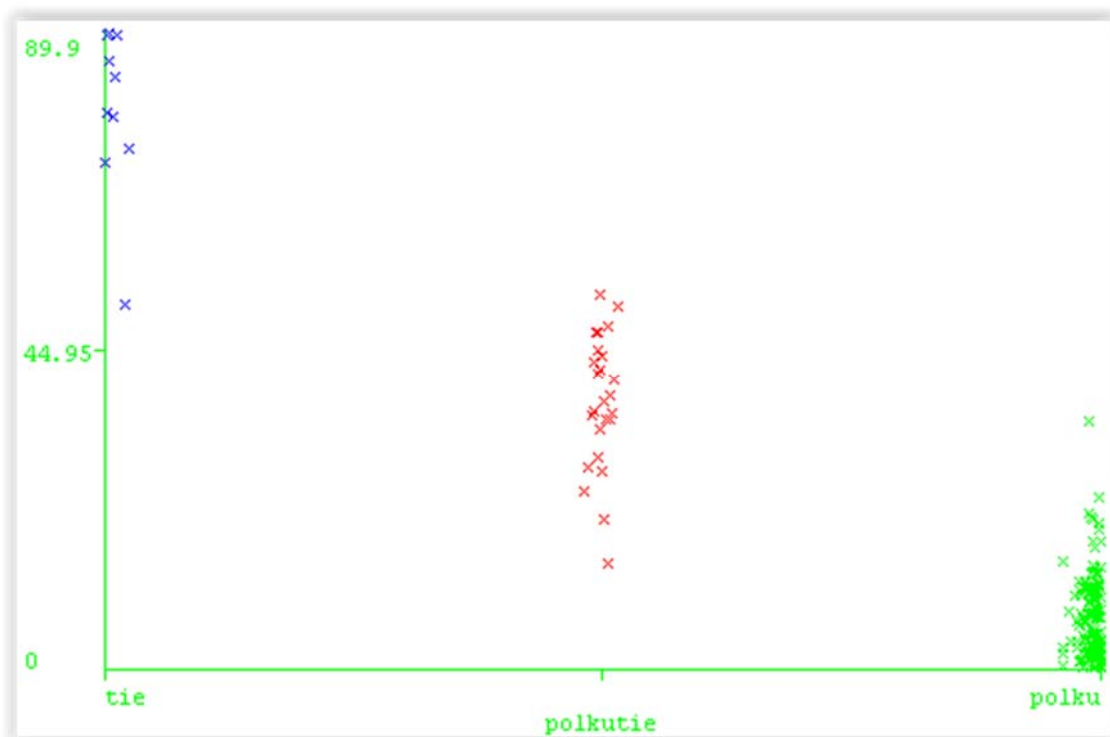
## 4.3 Segmentointi

Segmentoinnilla reitti voidaan jaotella erilaisiin liikkumisen muotoihin, kuten autoilun, pyöräilyyn, jalan kulkemiseen tai taksilla ajoon. GPS-piste sisältää x- ja y-koordinaatin, sekä aikaleiman. GPS-reitti on joukko GPS-pisteitä ajallisesti peräkkäisinä. Zheng ja muut [2010] olettavat, että kahden eri liikkumistavan välillä pitää olla GPS-piste tai lyhyt segmentti, jossa nopeus ja kiihtyvyys ovat vähäisiä. Esimerkiksi kulkutavan muutoksessa, siirryttäessä autolla ajosta linja-autolla matkustamiseen, täytyy kävellä autolta pysäkille. Kävely segmenttien tunnistaminen GPS-reitistä on siis ensimmäinen vaihe segmenttien erottamisessa toisistaan ja käveleminen toimii liikkumistavan vaihdon tunnisteena. Muissa segmenteissä sijaitsevien hitaiden osuuksien luokittelun estämiseksi kävelysegmenteiksi, esimerkiksi ruuhkassa ajamisen, käytetään segmentin pituutta, jolle on määritetty kynnyksarvo, esimerkiksi 200 m, jonka ylittyessä se luokitellaan ei-kävely segmentiksi.

### 4.3.1 Nopeus

Maastoajossa nopeus on suurin segmenttejä erottava tekijä. Tutkielmaa varten kerätystä aineistosta voidaan nopeuden avulla J48-luokittelijalla erotella polulla ajetuksi yhdeksän paikkatietoa kymmenestä, 23 paikkatietoa 25:stä polkutiellä ajetuksi ja 160 paikkatietoa 164:stä tiellä ajetuksi.

Rajat luokkien välillä ovat 24.4 km/h ja 55.5 km/h, joiden mukaan polulla ajaminen sijoittuu alle 24.4 kilometrin tuntinopeuksiin, tiellä ajaminen yli 55.5 kilometrin tuntinopeuksiin ja polkutiellä ajaminen näiden luokkien väliin. Luokkien välinen nopeusjakauma on esitetty kuvassa 4.



Kuva 4. Luokkien välisten nopeuksien jakauma maastoajossa.

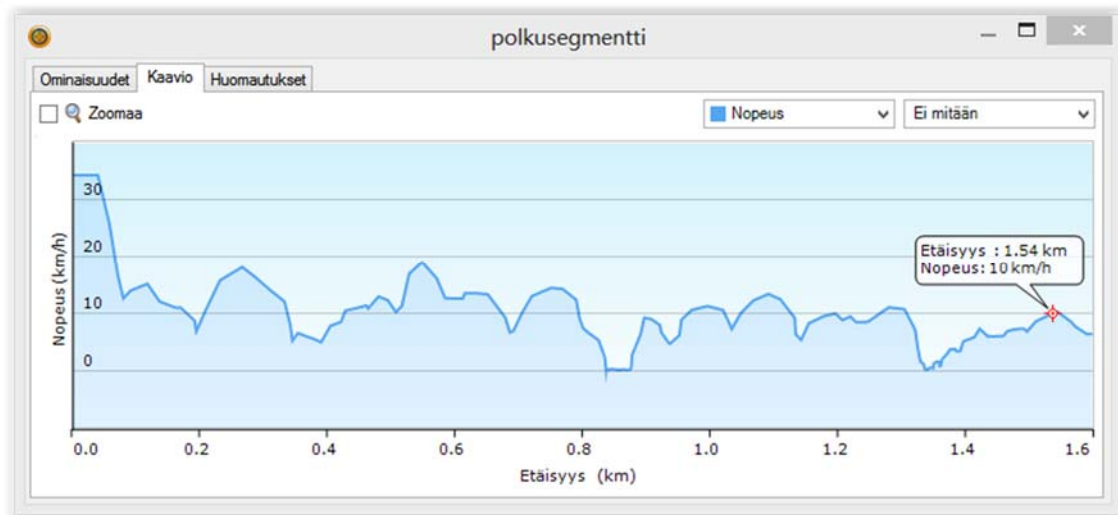
### 4.3.2 Nopeuden vaihtelu

Nopeuden vaihtelu lasketaan siten, että nopeuden muutos mitataan kahden peräkkäisen reittipisteen nopeuden prosentuaalisena muutoksena suhteessa alkunopeuteen kaavalla  $|(N_x - N_{x-1})/N_{x-1}| \cdot 100$ , jossa  $N_x$  on reittipisteen nopeus ja  $N_{x-1}$  sitä edeltävän reittipisteen nopeus.

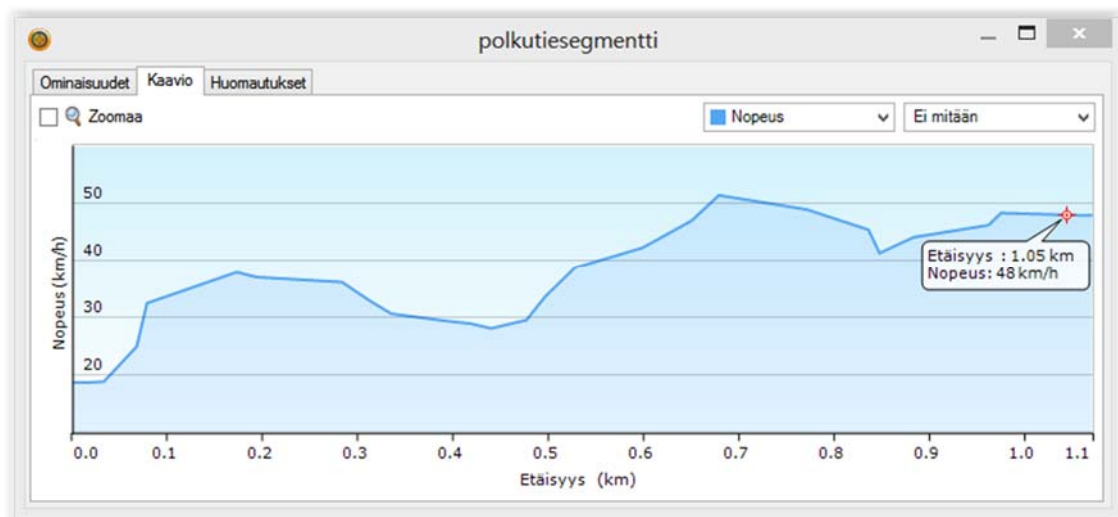
Nopeuden suhteellisen muutoksen avulla voidaan myös erotella erilaisia kulureittejä, kuten polku, polkutie tai isompi tie maastomoottoripyörällä ajattaessa. Polulla kiihdytettäessä vauhtia ja jarrutettaessa, pienestä keskituntinopeudesta johtuen, suhteelliset muutokset ovat isompia kuin leveämmällä ja vauhdikkaammalla metsätiellä. Vastaavasti metsätiellä kiihdytyksien ja jarrutuksien



suhteelliset muutokset nopeuden suhteen ovat isompia kuin tiellä. Kuvissa 5, 6, ja 7 on havainnollistettu eri segmenttien nopeuksia ja nopeuksien vaihtelua tutkielmaa varten kerätystä aineistosta. Huomio kiinnittyy nopeuden tasaisuuteen tiesegmentillä ja vastaavasti polkusegmentin nopeuden epätasaisuuteen suhteessa ajettuun matkaan. Polkutien nopeusvaihtelu on jossain näiden kahden välimaastossa.

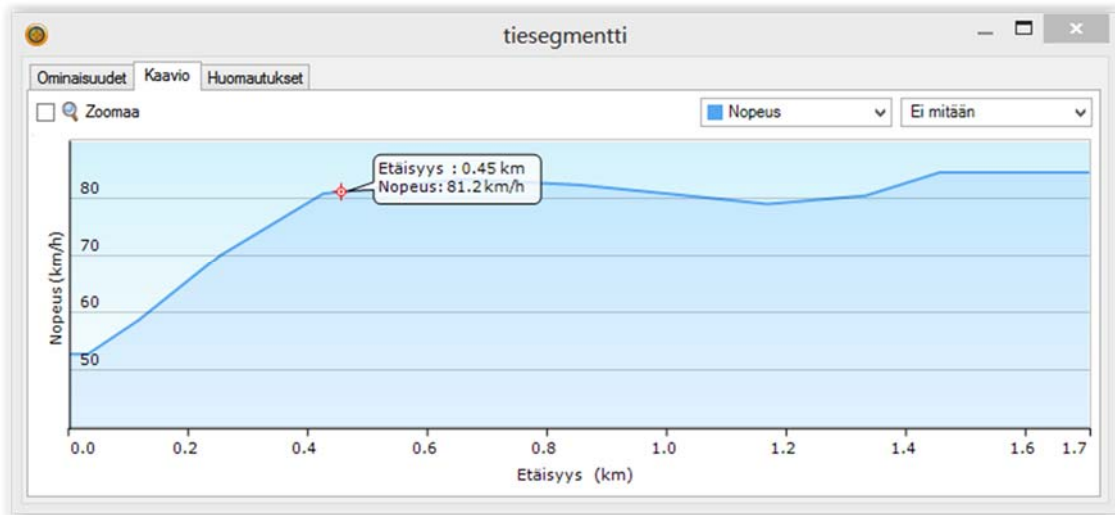


Kuva 5. Nopeus polulla ajettaessa.



Kuva 6. Nopeus polkutiellä ajettaessa.

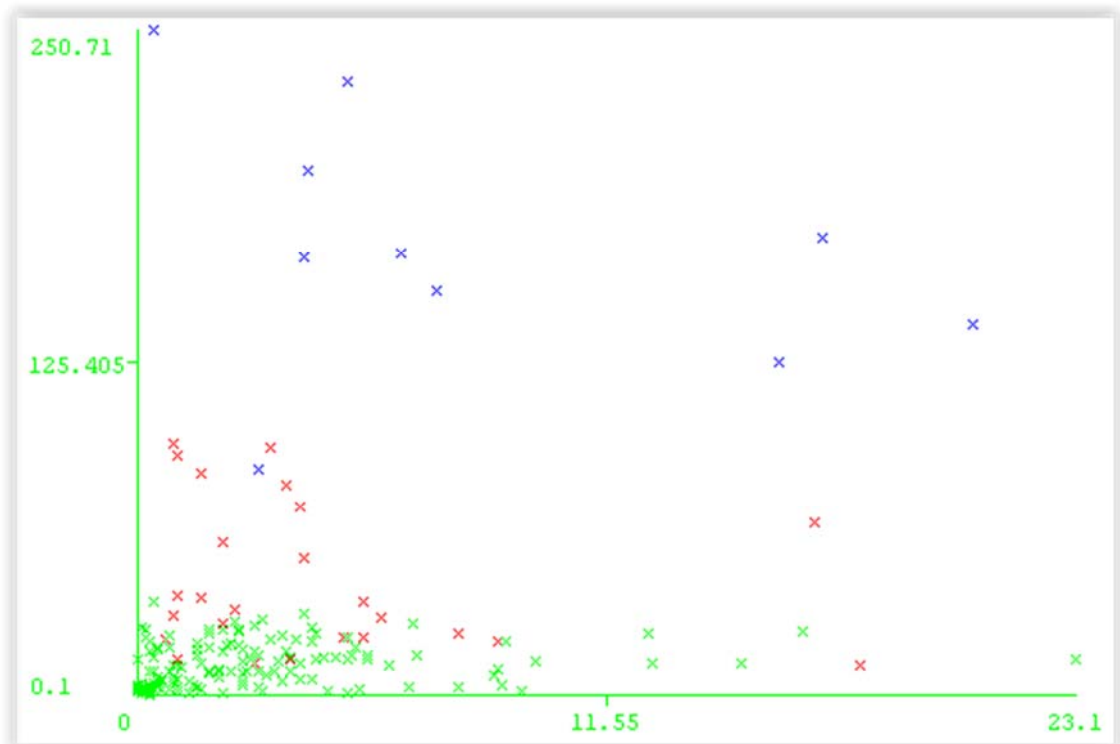




Kuva 7. Nopeus tiellä ajettaessa.

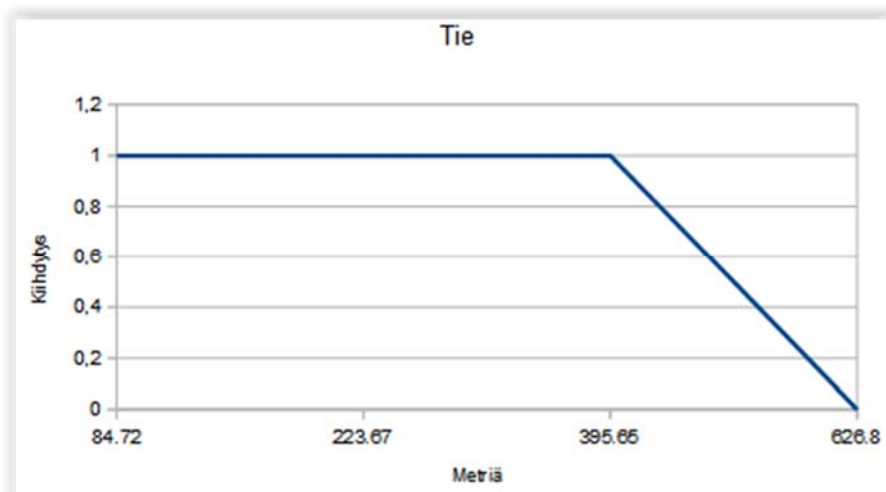
Zheng ja muut [2010] ovat löytäneet eri liikkumistavoissa eroja ainakin linja-auton, autoilun ja kävelyn osalta. Liikkeen pysähtymispisteiden tiheydellä voidaan erotella esimerkiksi linja-autolla ajo kaupunkiolosuhteissa, autolla ajo ja kävely. Käveltäessä pysähdytään enemmän kuin ajettaessa linja autolla ja autolla ajettaessa kaikkein harvemmin. Kaupunkiympäristössä linja-autot pysähtelevät useammin kuin autot ja kävelijät useammin kuin linja-autot.

Maastoajossa jarrutusten ja kiihdytysten esiintymistiheyttä ajan ja matkan suhteen tukee nopeuden muutosten ja GPS-laitteen rekisteröimien pisteiden välimatkan sirontakuvio kuvassa 8. Pysähdyksiä ei tutkimuksen aineistoon ole otettu. Voidaan kuitenkin olettaa, että ainakin vaikeammassa maastossa ajettaessa pysähdyksiä tulee ja tätä voitaisiin käyttää erilaisten reittien luokittelussa hyväksi. Paljon riippuu tietysti ajajan taitotasosta, miksei kunnostakin.

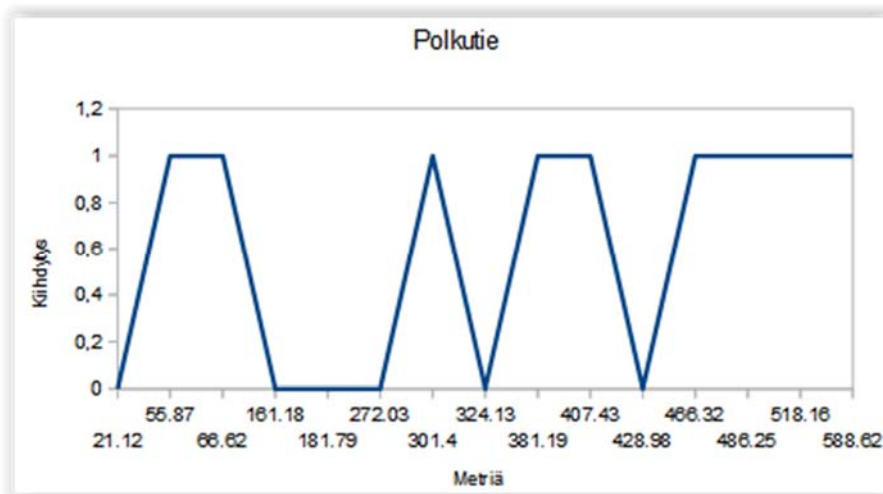


Kuva 8. Nopeuden muutokset GPS-laitteen rekisteröimien pisteiden välimatkojen suhteen.

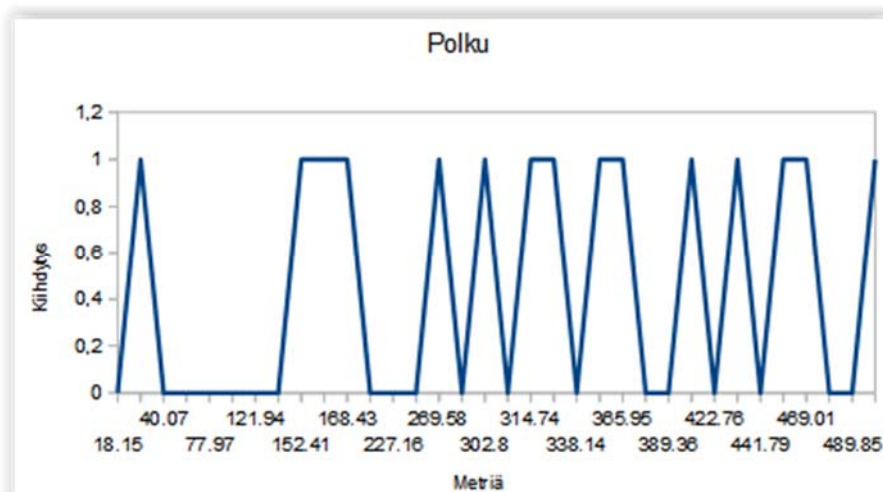
Pysähtymisiä havainnoivaa menetelmää [Li et al., 2011], jossa reittipisteelle, jossa ollaan paikallaan annetaan arvo 0, ja reittipisteelle, jossa ollaan liikkeellä, annetaan arvo 1, voidaan soveltaa myös jarrutuksiin ja kiihdytyksiin. Menetelmän etuna on hälyn poistuminen datasta, jolloin datan ominaisuudet on helppo havaita. Kuvissa 9, 10 ja 11 on tutkimusta varten tallennetusta reittitiedosta havainnollistettu noin puolen kilometrin matkalta kiihdytysten ja jarrutusten osuus tietynlaisessa maastossa ajettaessa.



Kuva 9. Binäärisekvenssi kiihdytyksistä tiellä.



Kuva 10. Binäärisekvenssi kiihdytyksistä polkutiellä.



Kuva 11. Binäärisekvenssi kiihdytyksistä polulla.

#### 4.3.3 Suunnan vaihtuminen

Autolla ajaessa tai yleensä tiellä liikuttaessa liikutaan tien mukaan ja esteitä ei ole. Näin ollen suunnan vaihdoksia on melko vähän verrattaessa metsässä tai polulla kävelyyn. Käveltäessä metsässä suunnanvaihdoksia on huomattavasti enemmän, johtuen polkujen muotoutumisesta suunnittelematta maastonmuotojen mukaan.

Suunnan vaihtumisen taajuutta voidaan käyttää segmentin tunnistuksessa. Taajuus määritellään esimerkiksi jakamalla segmentin niiden GPS-pisteiden lukumäärä, joissa suunta muuttuu, segmentin pituudella [Zheng et al., 2010]. Suunnan vaihtumisen taajuutta laskettaessa ongelmia muodostaa automaattisen

tietojenkäsittelyn kannalta kurssin muuttuminen asteikon yli esimerkiksi kolmesataaviisikymmentäkahdeksasta kahteen asteeseen. Myös jonkinlainen kynnysarvo kurssimuutoksen rekisteröimiselle täytyy asettaa, jotteivat pienet, ehkä tallentavasta laitteistostakin johtuvat muutaman asteen kurssimuutokset aiheuta väärää tulkintaa. Nämä kaksi asiaa tulee siis ottaa huomioon esikäsittelyvaiheessa. Suunnan vaihtumisen taajuuden avulla voidaan erotella toisistaan erilaisia kulkemisen tapoja, kuten autolla ajaminen, kävely tai polkupyöräily. Mahdollisesti myös yhden kulkutavan sisällä voidaan erottaa erilaisia kulkureittejä. Maastopyörällä ajettaessa kurssin muutokset korreloivat tiestön tai polun mutkitteluun. Poluthan ovat yleensä aika mutkaisia verrattuna teihin, ja mitä isompi tie on, sitä suurempi se yleensä on.

Tutkielmaa varten kerätyssä aineistossa kurssimuutoksilla havaittiin olevan pieni mahdollisuus eri maastokategorioiden luokitteluperusteeksi. Tiedon varmistaminen edellyttäisi eri paikoista sekä eri ajajien ajamana kerättyä laajempaa aineistoa ja jatkotutkimusta.

## 5. Yhteenveto

Tiedonlouhinnan prosesseilla paikantimen tallentamasta reitistä voidaan louhia lisää käyttäjien kiinnostuksen kohteita ja reitin luonteesta lisäinformaatiota. Paikanninlaitteet tallentavat pienelläkin matkalla suuren määrän numeerista tietoa, jonka käsittelyyn tiedonlouhinta soveltuu erinomaisen hyvin. Erilaisia mahdollisuuksia reittipalvelujen reittien esitystavan kehittämisessä on paljon.

Pysähtymisiä havainnoivaa menetelmää [Li et al., 2011], jossa reittitieto muutetaan binäärikoodiksi, voidaan soveltaa moneen muuhunkin tarkoitukseen, kuten maastoajossa erilaisten reitistä eristettyjen segmenttien havainnoimiseen. Nopeuden ohella nopeuden vaihtelut matkan ja ajan suhteen sekä kurssimuutoksien vastaavat vaihtelut osoittivat lupaavia mahdollisuuksia reittitiedon analysoimisessa. Asian tutkiminen laajemmin edellyttäisi ainakin tutkimukseen tarkoitetun aineiston tallentamista useilla erilaisilla GPS-tallentimilla. Tässä tutkielmassa käytetyn laitteen ominaisuudet, esimerkiksi paikan tallentaminen ajallisesti tietyin väliajoin ja sen aiheuttama mahdollinen reittitiedon vääristymä, jäivät arveluttamaan. Kovassa vauhdissa laite rekisteröi paikan, matkan suhteen, harvemmin kuin hitaassa.

## Viiteluettelo

[http://en.wikipedia.org/wiki/Point\\_of\\_interest](http://en.wikipedia.org/wiki/Point_of_interest)

[Kallioniemi, 2012] Pekka Kallioniemi, Maamerkit mobiilisuunnistuksen apuna. Pro gradu -tutkielma, Informaatiotieteiden yksikkö, Tampereen yliopisto, 2012.

[Li et al., 2011] Zhenhui Li, Jiawei Han, Ming Ji, Lu-An Tang, Yintao Yu, Bolin Ding and Roland Kays, MoveMine: Mining moving object data for discovery of animal movement patterns. *ACM Transactions on Intelligent Systems and Technology* 2, 4, Article 37, (2011).

[Zheng et al., 2010] Vincent W. Zheng, Yu Zheng, Xing Xie and Qiang Yang, Collaborative location and activity recommendations with GPS history data. In: *Proc. of the 19th International Conference on World Wide Web* (2010), pages 1029-1038.

[Zheng et al., 2010] Yu Zheng, Yukun Chen, Quannan Li, Xing Xie and Wei-Ying Ma, Understanding Transportation Modes Based on GPS Data for Web Applications. *ACM Transactions on The Web* 4, Article 1, 2010.

# Vaatusmäärittelyn rooli yhdistyvien organisaatioiden tietojärjestelmien integraatiossa

**Simo Pönni**

## Tiivistelmä.

Tutkielman tarkoitus on selvittää, millä tavalla ohjelmisto- ja tietojärjestelmäkehityksen vaatimusmäärittelyn menetelmiä ja käsitteistöä on esitelty alan tieteellisessä kirjallisuudessa, kun menetelmien sovelluskohde on kahden organisaation tietojärjestelmien ja IT:n integraatio.

**Avainsanat ja -sanonnat:** vaatimusmäärittely, tietojärjestelmät, integraatio

## 1. Johdanto

Nykyaikaisen tietojärjestelmän voi karkeimmillaan kuvailla ihmisistä ja tietokoneista koostuvana toiminnallisena kokonaisuutena, joka käsittelee tietoa. Tietoyhteiskunnassa tietojärjestelmät ovat varmasti tavalla tai toisella osa jokaisen yksityishenkilön arkipäiväistä elämää. Sitäkin varmemmin ne ohjaavat tai ovat läsnä helpottamassa jokaisen tavoitteellisen organisaation toimintaa, oli kyseessä sitten yritys, yhdistys, kunta tai kaupunki.

Kun organisaatiot yhdistyvät, esimerkiksi yrityskaupan seurauksena, on todennäköistä, että yhteistoiminnan mahdollistamiseksi joko olemassa olevien järjestelmien on voitava olla vuorovaikutuksessa keskenään tai rinnakkaisista järjestelmistä kannattaa luopua. Kun kahden yhdistyvän organisaation toiminta on riittävän samankaltaista keskenään, rinnakkaisista järjestelmistä luopuminen voi olla organisaatioiden yhdistymisen lähtökohta. Samankaltaista toimintaa tukevan laitteiston hankinta ja ylläpito sekä sisäisten palveluiden tuottaminen on edullisempaa, kun se tehdään keskitetysti. Organisaatioiden tietojärjestelmien ja IT:n integraatio ei siis välttämättä ole ainoastaan organisaatioiden yhdistymisen seuraus, vaan se voi olla itseisarvoisesti tärkeä asia.

Organisaatioiden yhdistymistä – erityisesti yritysfuusioiden ja yritysostojen näkökulmasta – on tutkittu vuosikymmenten ajan, mutta siihen liittyvä tietojärjestelmien ja informaatioteknologian integraatio on viime vuosia lukuun ottamatta saanut vain vähän huomiota osakseen [Lange, 2015].

Tietojärjestelmien ja IT:n integraation läpiviennissä on epäilemättä mahdollista ja hyödyllistä soveltaa erilaisia ohjelmisto- ja tietojärjestelmäkehityksen vaatimusmäärittelyn menetelmiä. Tämän tutkielman tarkoitus on selvittää, millä tavalla vaatimusmäärittelyä on tarkasteltu tieteellisessä kirjallisuudessa organisaatioiden yhdistymisen ja järjestelmäintegraation viitekehyksessä.

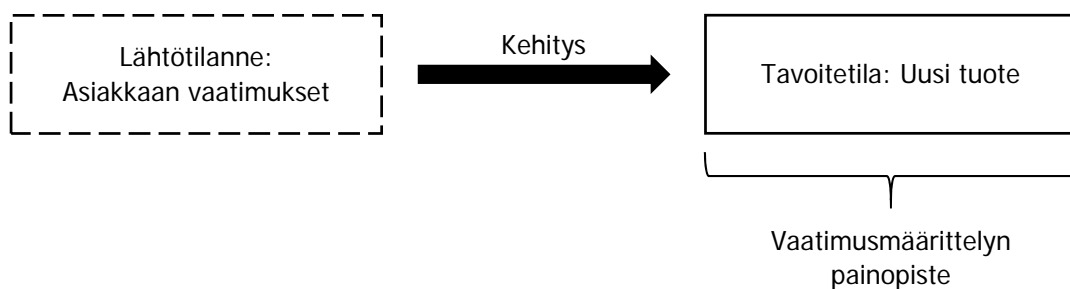
Tutkielman luku 2 esittelee lyhyesti tutkielmassa käsiteltävät aihealueet sekä taustan ja motivaation tutkimuskysymysten asettelulle luvussa 3. Luvussa 4 käydään läpi, kuinka kirjallisuushaku suoritettiin ja esitellään löydökset. Lopuksi viidennessä luvussa arvioidaan, kohtaavatko organisaatioiden yhdistymistä ja vaatimusmäärittelyä käsittelevä tutkimus toisensa, sekä pohditaan, miltä osin näiden kahden tutkimusalan välille olisi kannattavaa rakentaa yhteyksiä.

## 2. Vaatimusmäärittely ohjelmisto- ja järjestelmäkehityksessä

Ohjelmisto- ja järjestelmäkehityksen vaatimusmäärittely on oma tutkimusalueensa, joka keskittyy vaatimusmäärittelyn haasteiden tutkimiseen. Tutkimuksen tarkoitus on tuottaa tehokkaita vaatimusten nostattamisen, arvioinnin ja dokumentoinnin malleja sekä keinoja hallinnoida kehitysprojektien aikana muuttuvia vaatimuksia [Sommerville, 2007].

Ohjelmistojen ja tietojärjestelmien kehitys edellyttää aina tuotteen vaatimusten määrittelyä ja dokumentointia jollakin tasolla, vaikka dokumentoinnin määrä ja laatu voi vaihdella. Perinteisiä ohjelmistokehityksen malleja noudattavissa projekteissa vaatimusmäärittely ja dokumentointi pyritään ideaalitilanteessa tekemään valmiiksi ja rakentamaan tilattu tuote peilaamalla valmiiseen dokumentaatioon. Ketterissä, iteratiivisissa kehitysmalleissa pyritään välttämään vaatimusten raskasta dokumentointia ennalta. Tällöin keskeneräisestä tuotteesta tulee osa jatkuvasti muuttuvaa dokumentaatiota, jonka pohjalta tehdään pienissä osissa uutta vaatimusmäärittelyä.

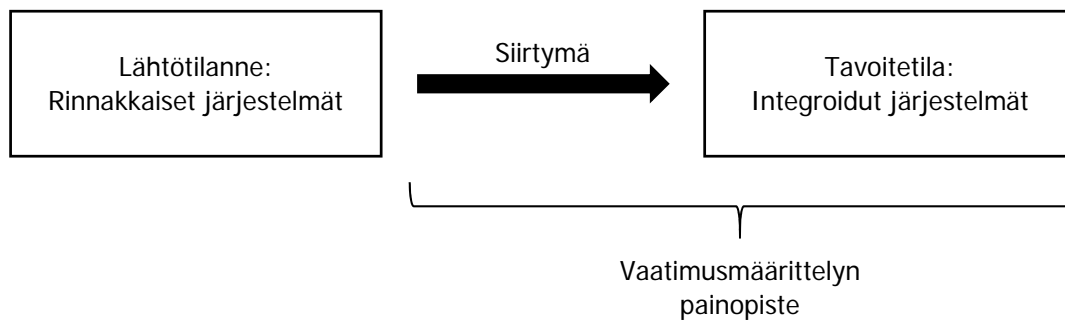
Uutta tuotetta kehitettäessä on mahdollista, että tuotteen vaatimukset voivat asettaa rajoitteita toteutustavalle, esimerkiksi määrittelemällä käytettävän ohjelmointikielen, mutta vaatimusmäärittelyn merkitys itse kehitysvaiheen säätelyssä on vähäinen. Sen ensisijainen tavoite on määritellä ja kommunikoida kehitettävän tuotteen ominaisuudet tuotteen toteuttajille projektin asiakkaan vaatimusten mukaisesti (ks. kuva 1).



Kuva 1: Vaatimusmäärittelyn painopiste ohjelmistokehityksessä

Tutustuminen vaatimusmäärittelyn alan kirjallisuuteen paljastaa pian, että useimmiten kirjoittaja implisiittisesti olettaa menetelmiä sovellettavan nimenomaisesti uuden tuotteen kehitysprosessissa. Tällöin kehitysvaiheen itsensä säätelyn merkitys jää vähäiseksi vaatimusmäärittelyn näkökulmasta ja korostuu esimerkiksi ohjelmisto- tai järjestelmäprojektin johtamisen osa-alueena.

Järjestelmäintegraation tyypilliset haasteet voivat poiketa uuden järjestelmän kehittämisen haasteista tältä osin. On mahdollista, että taloudellisista tai aikataullisista syistä uusia toiminnallisia ratkaisuja ei ole kannattavaa kehittää, vaan yksi organisaatio joutuu sovittamaan prosessinsa toisen organisaation järjestelmäratkaisujen malliin (takeover-integraatio). Lisäksi on todennäköistä, että integraation missä tahansa skenaariossa lähtökohtana on tilanne, jossa olemassa olevilla järjestelmillä on käyttäjiä, eikä työskentely saa katketa. Tästä seuraa, että siirtymävaihetta on voitava säädellä ja itse siirtymälle on voitava asettaa vaatimuksia.



Kuva 2: Vaatimusmäärittelyn painopiste järjestelmäintegraatiossa

Tämän tutkielman oletus on, että ohjelmisto- ja järjestelmäkehityksen vaatimusmäärittelyn menetelmät ovat sovellettavissa myös järjestelmäintegraatioon. Antamalla enemmän painoarvoa organisaatioiden välisen integraation keskeisille käsitteille ja aihealueille, kuten organisaatorakenteille, prosesseille, sidosryhmille ja sidosryhmäkohtaisille näkökulmille, sekä nostamalla siirtymä keskeiseksi vaatimusmäärittelyn kohteeksi (ks. kuva 2), useat menetelmät ovat sovellettavissa sellaisenaan.

Tutkielman pyrkimys on selvittää, onko tieteellisessä kirjallisuudessa ohjelmisto- ja järjestelmäkehityksen vaatimusmäärittelyn menetelmiä ehdotettu sovellettavaksi organisaatioiden yhdistymisestä seuraavan järjestelmäintegraation kontekstissa.



### 3. Tutkimuskysymysten asettelu

Taulukossa 1 esitellään tiivistetysti tutkielman keskeiset kysymykset, jotka keskittyvät vaatimusmäärittelyä, tietojärjestelmien integraatiota ja organisaatioiden yhdistymistä tutkivien alojen nykytilaan.

#	Tutkimuskysymys	Perustelu / Tarkennus
K1	Onko tehty tieteellistä tutkimusta, jossa vaatimusmäärittelyn rooli organisaatioiden tietojärjestelmien integraatiossa olisi ollut tarkastelun kohteena?	Tämä kysymys on tutkielman lähtökohta, jota seuraavat kysymykset tämentävät.
K2	Kuinka paljon vaatimusmäärittelyn ja järjestelmäintegraatiot yhdistävää tieteellistä kirjallisuutta löytyy?	Vaikka kysymyksen K1 aihealuetta erityisesti käsittelevää kirjallisuutta ei löytyisikään, on kiinnostavaa tietää, onko aihetta sivuttu vain vaatimusmäärittelyn ja järjestelmäintegraatiot yhdistävissä artikkeleissa.
K3	Kuinka paljon järjestelmäintegraatioita ja organisaatioiden yhdistymistä käsittelevää tieteellistä kirjallisuutta löytyy?	Kysymyksen K2 perustelua vastaavasti voi olla mahdollista, että K1:n aihetta on sivuttu vain organisaatioiden yhdistymisen ja järjestelmäintegraatiot yhdistävissä artikkeleissa.
K4	<b>Jatkokysymys:</b> Millä tavalla vaatimusmäärittelyä ja organisaatioiden yhdistymistä tutkivien alojen julkaisut kohtaavat toisensa, kun niissä käsitellään yhdistävänä tekijänä myös tietojärjestelmien integraatiota?	Jos oletetaan, että vaatimusmäärittelyn ja organisaatioiden yhdistymistä tutkivien alojen välille on hyödyllistä tai tarpeellista rakentaa siltoja järjestelmäintegraatioita ajatellen, on tarpeellista tietää, mistä näkökulmasta kannattaa rakentaa. Onko vaatimusmäärittelyn menettelyjen hyödyllisyyteen mahdollisesti aiemmin viitattu organisaatioiden yhdistymistä käsittelevissä teksteissä? Tai onko organisaatioiden yhdistyminen esitetty selvänä kontekstina vaatimusmäärittelyn menetelmien soveltamiselle järjestelmäintegraatitilanteessa? Kysymyksistä K2 ja K3 seuraavien kirjallisuushakujen tulosten tarkempi analysointi vastaa näihin kysymyksen.

*Taulukko 1: Tutkimuskysymykset*

Taulukossa 1 esiteltuihin kysymyksiin voidaan vastata tekemällä hyvin määritelty kirjallisuuskatsaus. Tutkielman aikataulu asettaa rajoitteita kirjallisuuskatsauksen laajuudelle: systemaattiseen kirjallisuushakuun ja sen ammattimaiseen

analyysiin, sekä tulosten yksityiskohtaiseen erittelyyn voisi mennä henkilötyövuosia. Edellä mainitusta syystä tutkielma keskittyy ensisijaisesti vastaamaan keskeisiin kysymyksiin K1-K3. Viimeisessä luvussa pohditaan jatkokysymystä K4 ja mihin mahdollisessa jatkotutkimuksessa kannattaisi keskittyä.

## 4. Kirjallisuuskatsaus

Tutkielman kirjallisuuskatsausosuus on tehty ainoastaan digitaalisia, pääasiassa englanninkielisiä tekstejä sisältäviä lähdetietokantoja hyväksi käyttäen. Tässä luvussa esitellään lähteet, joista tietoa on haettu ja kuinka niiden käyttöön on päädytty. Lisäksi esitellään kirjallisuuskatsauksen hakutermit ja muodostetut hakulauseet, sekä lauseiden muodostusperusteet. Lopuksi käydään läpi haun tulokset ja perusteet, joilla tulos hyväksyttiin tai hylättiin.

### 4.1. Lähteet

Tammikuussa 2015 julkaistiin laajamittainen kirjallisuuskatsaus ontologioiden käytöstä vaatimusmäärittelyssä [Dermeval et al., 2015]. Katsaus on tuore ja liittyy läheisesti tutkielman aihealueeseen, joten voidaan olettaa siinä käytettyjen tietokantojen olevan käyttökelpoisia lähteitä.

Katsauksen seitsemästä lähteestä tutkielmaan valikoitui neljä. Valintaperuste oli mahdollisuus tietokannan artikkeleiden laajaan käyttöön Tampereen yliopiston yhteisötunnistautumisen kautta.

#	Lähdetietokanta	URL
L1	ACM Digital Library	<a href="http://dl.acm.org/">http://dl.acm.org/</a>
L2	IEEE Xplore	<a href="http://ieeexplore.ieee.org/">http://ieeexplore.ieee.org/</a>
L3	ScienceDirect	<a href="http://www.sciencedirect.com/">http://www.sciencedirect.com/</a>
L4	SpringerLink	<a href="http://link.springer.com/">http://link.springer.com/</a>

*Taulukko 2: Lähdetietokannat*

Kaikki taulukon 2 tietokannat tarjoavat jonkinlaisen mahdollisuuden ehtolausekeisiin perustuvan hakujen tekemiseen.

### 4.2. Hakulausekkeet ja tulosjoukot

Suurin osa vaatimusmäärittelyn tutkimusalan tieteellisestä tekstistä on englanninkielistä, joten haku tehtiin käyttäen englanninkielisiä hakusanoja. Tutkimuskysymyksen K1 kannalta keskeisimmät aihealueet ja niitä vastaavat englanninkieliset hakulausekkeet on listattu taulukossa 3.

Jotta haun perimmäisenä kohteena oleva joukko voitaisiin esitellä ytimekkäästi, määritellään tutkielmassa käsiteltävät haun tulosten osajoukot.

## Määritelmä

Taulukon 3 rivi määrittelee joukon, joka on yhdistelmä lähdetietokantojen hakutuloksena palauttamien julkaisujen joukoista, kun tehtyihin hakuihin pätevät seuraavat ehdot:

1. Haut on suoritettu kaikissa taulukon 2 tietokannoissa ja vain taulukossa 2 esiintyvissä tietokannoissa.
2. Haut on suoritettu käyttäen taulukossa 3 joukon määrittelevällä rivillä hakulauseke-sarakkeessa esitettyjä lausekkeita.
3. Haut on kohdistettu käytetyn lähdetietokannan oletuskohdistusten mukaisesti (kaikissa taulukon 2 tietokannoissa ainakin julkaisun otsikkoon, tiivistelmään ja kokotekstiin).
4. Muita hakuehtoja ei ole asetettu.

#	Joukko	Aihealue	Hakulauseke
J1	RE	Vaatimusmäärittely	("requirements engineering" OR "requirement engineering")
J2	SE	Järjestelmäsuunnittelu, järjestelmäkehitys	("systems engineering" OR "system engineering")
J3	IS	Tietojärjestelmät	("information system" OR "information systems")
J4	IT	Informaatioteknologia	("information technology" OR "IT")
J5	IN	Integraatio	("integration" OR "integrations")
J6	MA	Yritysfuusiot ja -ostot, organisaatioiden yhdistyminen	("merger" OR "mergers" OR "acquisition" OR "acquisitions")

*Taulukko 3: Tulosten osajoukot ja niitä vastaavat hakulausekkeet*

Jotta tutkimuskysymykseen K1 voitaisiin vastata, on etsittävä julkaisuja, jotka käsittelevät vaatimusmäärittelyn aihealuetta. Tehdyissä hauissa vaatimusmäärittelyn ja järjestelmäkehityksen aihealueita voidaan käsitellä yhdisteenä ( $RE \cup SE$ ), koska tutkielmassa ollaan kiinnostuneita erityisesti järjestelmistä ja vaatimusmäärittelyn katsotaan kiinteästi kuuluvan järjestelmäkehityksen osa-alueeksi [Jackson & Stevens, 2000].

Hakutuloksena kiinnostavan julkaisun tulee siis käsitellä tietojärjestelmiä. Koska informaatioteknologia on olennainen osa organisaatioiden tietojärjestelmiä, voidaan yhdistettävä  $IS \cup IT$  käsitellä yhtenä kokonaisuutena ja näin saada enemmän potentiaalisesti kiinnostavia tuloksia. Edellä mainittujen aihealueiden lisäksi julkaisun on käsiteltävä integraatiota (IN) organisaatioiden yhdistymisen kontekstissa (MA).

$$(RE \cup SE) \cap (IS \cup IT) \cap IN \cap MA$$

*Lauseke 1: Tutkimuskysymyksen K1 kannalta kiinnostava joukko*

Julkaisujen joukko, jonka perusteella tutkimuskysymykseen K1 pystytään vastaamaan, voidaan taulukossa 3 määriteltyjen osajoukkojen avulla lopulta esittää joukko-opillisena lausekkeena 1.

Lähdetietokantoihin tehtävät haut voidaan muodostaa korvaamalla joukko sen määrittelevällä hakulausekkeella, korvaamalla leikkausoperaattori sanalla "AND" ja yhdisteoperaattori sanalla "OR".

#### 4.3. Ensimmäinen hakukierros – Lauseke 1, kohdistamaton haku

Tutkielman kirjallisuushaun suurin haaste oli kyselyn muotoilu tehokkaasti ja tasapainoisesti. Kyselyssä pyrittiin välttämään hakutulosten massan paisumista liian suureksi, mutta kuitenkin pitämään relevantit tulokset mukana.

Tapa, jolla hakutulosten määrää pyrittiin kontrolloimaan, oli haun kohdistaminen vain julkaisujen metatietoihin. Koska käytetyt lähdetietokannat eivät tarjonneet keskenään suoraan vertailukelpoisia haun kohdistamisen mahdollisuuksia, tehtiin ensimmäisellä kierroksella lausekkeeseen 1 perustuen täysin kohdistamaton haku, jota rajattiin myöhemmin lähdetietokantojen tarjoamien mahdollisuuksien mukaan.

#	Lähdetietokanta	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	740	4,02 %
L2	IEEE Xplore	8419	45,74 %
L3	ScienceDirect	4235	23,01 %
L4	SpringerLink	5012	27,23 %
	<b>Yhteensä</b>	<b>18406</b>	<b>100 %</b>

*Taulukko 4: Kohdistamattomien hakujen tulokset, lause 1*

Taulukon 4 yhteenvedosta nähdään, että tutkielman laajuudessa kaikkia artikkeleita ei olisi voitu käsitellä mielekkäällä tavalla. Kiinnostavin anti kohdistamattomassa haussa oli se, mikä potentiaali eri lähdetietokannoilla saattaisi olla tulevia hakuja tehdessä. Jotta saavutettiin tarkemman tarkastelun mahdollistava julkaisujen joukko, ensimmäisen kierroksen hakutulosta oli rajattava.

#### 4.4. Ensimmäinen hakukierros – Lauseke 1, hakutulosten raja

Lähdetietokannat L1 ja L3 tarjosivat mahdollisuuden rajata hakutuloksia kohdistamalla alkuperäiset hakulausekkeet vain julkaisujen otsikkoihin ja tiivistelmiin. Tämä rajaustapa valittiin, koska voidaan olettaa, että otsikko ja tiivistelmä kiinnostavat julkaisun aihealueen, jolloin kohdistamalla haku erityisesti otsikkoon ja tiivistelmään, voidaan olettaa saatavan enemmän relevantteja tuloksia.

Tietokanta L2 tarjosi mahdollisuuden kohdistaa haku vain otsikkoon ja tiivistelmään, mutta kohdistuksen määrittelemisen eksplisiittisesti käytetyssä hakulauseessa kasvatti lauseen pituuden liian suureksi, eikä hakua voitu suorittaa. Rajausta päädyttiin tekemään kohdistamalla haku kaikkeen julkaisun metadataan, johon myös otsikko ja tiivistelmä sisältyvät.

Lähdetietokanta L4, SpringerLink, ei tarjonnut mahdollisuutta kohdistaa hakua, joten oli valittava erilainen lähestymistapa tulosten rajaukseen. SpringerLink antoi mahdollisuuden suodattaa hakutuloksia seuraavin kriteerein: Julkaisun tyyppi, Tieteenala (Discipline), Alahaara (Subdiscipline) ja Kieli. Lisäksi hakutuloksista olisi voitu jättää pois sisältö, josta on tarjolla ainoastaan esikatseluversio (Preview-Only Content), mutta näin ei tehty. Taulukko 5 esittää, kuinka SpringerLinkin hakutulokset suodatettiin vaiheittain.

#	Vaihe	Tulokset (kpl)
1	Kohdistamaton hakulauseke (RE ∪ SE) ∩ (IS ∪ IT) ∩ IN ∩ MA	5012
2	Julkaisun kieli = Englanti	4886
3	Tiedonhaara = Tietojenkäsittelytieteet (Computer Science)	3159
4	Alahaara = Ohjelmistokehitys (SWE, Software Engineering)	971
5	Alahaara = Tietojärjestelmät ja sovellukset (Information Systems and applications)	410
6	Alahaara = Yritysten tietojärjestelmät (Business Information Systems)	121

Taulukko 5: Springerlinkin hakutulosten suodattaminen, lauseke 1

#### 4.5. Ensimmäinen hakukierros – Rajattujen tulosten relevanssin arviointi

Kun kohdassa 4.4. kuvaillut rajaukset kohdistamattomiin hakuihin oli suoritettu, jäi jäljelle yhteensä 154 tulosta. Taulukko 6 esittää tulosten jakauman.

#	Lähdetietokanta	Rajaustapa	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	Otsikko ja tiivistelmä	3	1,95 %
L2	IEEE Xplore	Kaikki metadata	27	17,53 %
L3	ScienceDirect	Otsikko ja tiivistelmä	3	1,95 %
L4	SpringerLink	Julkaisun alan rajaus	121	78,57 %

Taulukko 6: Rajatut hakutulokset, lauseke 1

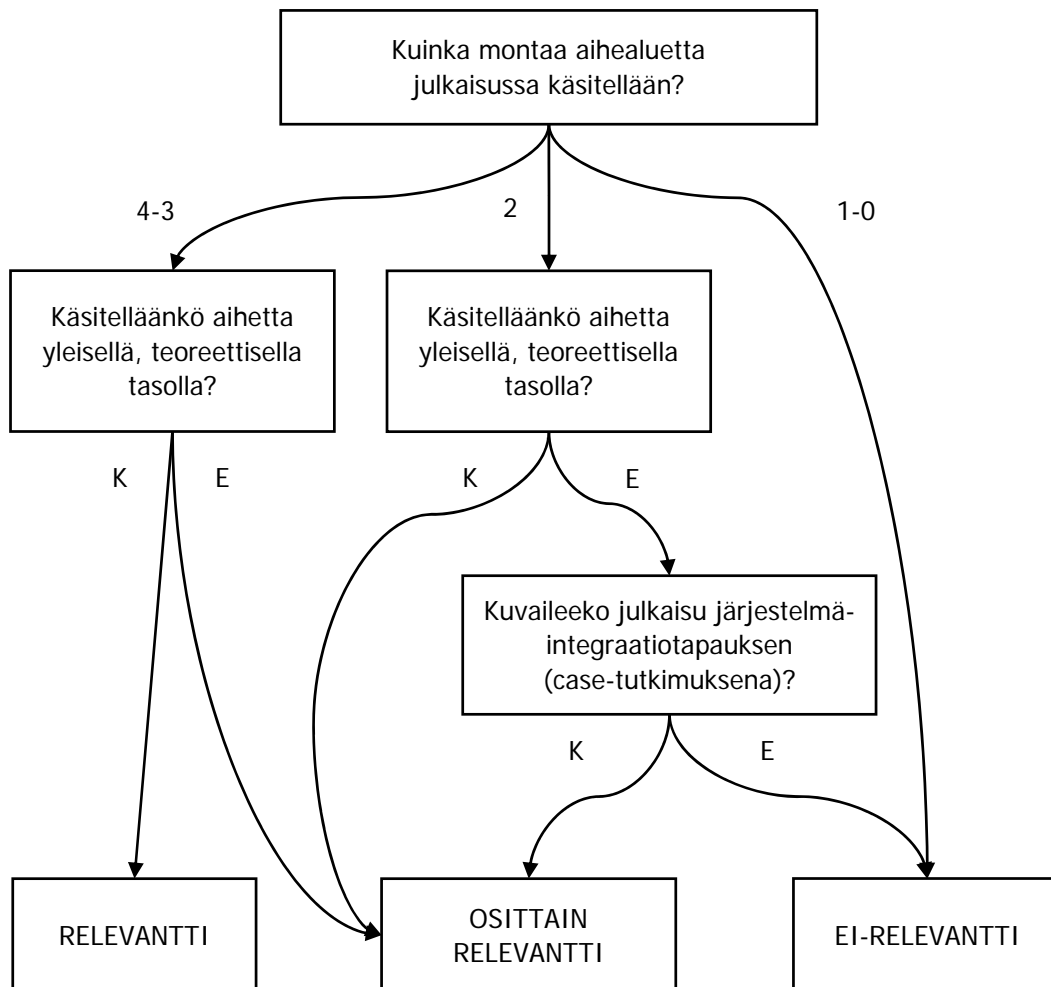
Tulosten relevanssin määrittämiseksi muotoiltiin yksinkertainen heuristiikka (ks. kuva 3), joka jakoi tulokset kolmeen kategoriaan: relevantit tulokset, osittain relevantit tulokset ja ei-relevantit tulokset.

Heuristiikan ensimmäisessä vaiheessa arvioitiin, kuinka monta kiinnostavaa aihealuetta löydetty teksti käsitteli. Aihealueita käsiteltiin taulukossa 7 esitetyinä kokonaisuuksina, jotka on muodostettu samoin perustein kuin lausekkeen 1 toisiaan leikkaavat osajoukot.

#	Aihealue	Joukko
1	Vaatimusmäärittely tietojärjestelmäkehityksessä	( RE $\cup$ SE )
2	Tietojärjestelmät	( IS $\cup$ IT )
3	Integraatio	IN
4	Yritysfuusiot ja -ostot, organisaatioiden yhdistyminen	MA

Taulukko 7: Tuloksen relevanssin arvioinnissa käytetyt osajoukot

Toisessa vaiheessa arvioitiin, onko julkaisun aihealueita lähestytty teoreettisella tasolla vai sovelluksen kautta, esimerkiksi case-tutkimuksena. Aiheen teoreettinen käsittely katsottiin arvokkaammaksi kirjallisuushaun kannalta.



Kuva 3: Tuloksen relevanssin arvioinnissa käytetty heuristiikka

Jos käsiteltäviä aihealueita on kaksi, oli mahdollisuus vielä kolmanteen vaiheeseen, jossa arvioitiin, esittelikö julkaisu järjestelmäintegraatiota koskevan case-tutkimuksen, jolloin se katsottiin osittain relevantiksi.

Julkaisujen relevanssin jako kolmeen on kiinnostavaa vain siksi, että se antaa yleiskatsauksen siihen, miten lähelle kysymyksen K1 aihealuetta julkaisut asetuvat. Kuitenkin vain relevantit tulokset analysoitiin tarkemmin ja vain niistä pohdittiin, edustivatko ne kysymyksessä K1 määriteltä tutkimusaluetta.

Arvioitujen 154 julkaisun joukossa oli vain yksi relevantiksi arvioitu tulos. Taulukko 8 esittelee tulosten relevanssin jakauman.

Tuloksen relevanssi	L1	L2	L3	L4	Yhteensä	Prosenttijakauma
Relevantti	0	1	0	0	1	0,65 %
Osittain Relevantti	1	1	1	5	8	5,19 %
Ei-Relevantti	2	25	2	116	145	94,16 %
<b>Yhteensä</b>	<b>3</b>	<b>27</b>	<b>3</b>	<b>121</b>	<b>154</b>	<b>100,00 %</b>

*Taulukko 8: Tulosten jakautuminen relevanssin ja lähdetietokannan mukaan*

#### 4.6. Toinen hakukierros - Lauseke 2

Toisen ja kolmannen hakukierroksen tavoitteena oli saada yleiskatsaus tutkielman kysymyksen K1 aihealuetta reunustaviin julkaisuihin. Haut suoritettiin samoissa lähdetietokannoissa kuin ensimmäisellä kierroksella. Hakujen kohdistus- ja rajausmenettelyt suoritettiin samoin kuin ensimmäisellä kierroksella, kohdassa 4.4. määritellyllä tavalla.

Lauseke 2 määrittelee joukon, jonka perusteella kysymykseen K2 voidaan vastata.

$$(RE \cup SE) \cap (IS \cup IT) \cap IN$$

*Lauseke 2: Tutkimuskysymyksen K2 kannalta kiinnostava joukko*

Hakuehtojen löysentäminen kasvatti tulosten määrän moninkertaiseksi verrattuna ensimmäiseen kierrokseen (ks. taulukko 9).

#	Lähdetietokanta	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	4272	5,10 %
L2	IEEE Xplore	36895	44,00 %
L3	ScienceDirect	19423	23,17 %
L4	SpringerLink	23254	27,73 %
	<b>Yhteensä</b>	<b>83844</b>	<b>100 %</b>

*Taulukko 9: Kohdistamattomien hakujen tulokset, lauseke 2*

#	Lähdetietokanta	Rajaustapa	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	Otsikko ja tiivistelmä	30	2,12 %
L2	IEEE Xplore	Kaikki metadata	695	49,19 %
L3	ScienceDirect	Otsikko ja tiivistelmä	92	6,51 %
L4	SpringerLink	Julkaisun alan rajaus	596	42,18 %

Taulukko 10: Rajatut hakutulokset, lauseke 2

Tulosten suuren määrän takia kattavaa läpikäyntiä ja relevanssiarviointia ei voitu tehdä, mutta hakutulosten otsikoista muodostui kokonaiskuva: vaatimusmäärittely (requirements engineering) ja tietojärjestelmien integraatio eivät esiintyneet samanaikaisesti keskeisinä aihealueina suurimmassa osassa tuloksia.

#### 4.7. Kolmas hakukierros - Lauseke 3

Kolmannella hakukierroksella pyrittiin vastaamaan kysymykseen K3, lausekkeen 3 kuvaileman joukon perusteella.

$$(IS \cup IT) \cap IN \cap MA$$

Lauseke 3: Tutkimuskysymyksen K3 kannalta kiinnostava joukko

Kohdistamattomien hakujen tuloksista (ks. taulukko 11) voidaan huomata, että yritysfuusioita ja -ostoja on tutkittu laajalti.

#	Lähdetietokanta	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	8024	2,16 %
L2	IEEE Xplore	81312	21,93 %
L3	ScienceDirect	164198	44,29 %
L4	SpringerLink	117235	31,62 %
	<b>Yhteensä</b>	<b>370769</b>	<b>100 %</b>

Taulukko 11: Kohdistamattomien hakujen tulokset, lauseke 3

Myös rajattu haku tuotti runsaasti tuloksia (ks. taulukko 12), jotka käytiin läpi vastaavalla tarkkuudella kuin toisen kierroksen rajatut hakutulokset. Vaikka vaatimusmäärittely aiheena ei esiintynyt tulosten otsikoissa millään tapaa, yritysfuusion jälkeistä tietojärjestelmien integraatiota käsittelevää aineistoa löytyi paljon. Tutkielman aiheen tarkempaa jatkokäsittelyä ajatellen tämä on haaste, mutta myös motiivi: jos vaatimusmäärittely ei esiinny tietojärjestelmien integraatiota käsittelevien tekstien osittaisena aihealueena organisaatioiden yhdistymisen kontekstissa ja voidaan todeta tutkimusalojen yhdistämisellä olevan arvoa, niin alojen yhdistämisestä saatava mahdollinen hyöty voi olla merkittävä.



#	Lähdetietokanta	Rajaustapa	Tulokset (kpl)	Prosenttiosuus
L1	ACM Digital Library	Otsikko ja tiivistelmä	52	2,39 %
L2	IEEE Xplore	Kaikki metadata	1119	51,50 %
L3	ScienceDirect	Otsikko ja tiivistelmä	578	26,60 %
L4	SpringerLink	Julkaisun alan rajaus	424	19,51 %

*Taulukko 12: Rajatut hakutulokset, lauseke 3*

## 5. Yhteenveto

Ensimmäisen hakukierroksen tulosten ja niiden relevanssianalyysin perusteella näyttää todennäköiseltä, että tutkimusta, joka käsittelee erityisesti vaatimusmäärittelyn roolia organisaatioiden yhdistymistä seuraavassa tietojärjestelmien integraatiossa, ei ole tehty. Tämä ei kuitenkaan välttämättä tarkoita sitä, että vaatimusmäärittelyn merkitykseen ei olisi viitattu esimerkiksi tutkimuksessa, joka tutkii organisaatioiden yhdistymisestä seuraavaa tietojärjestelmien integraatiota. Lisäksi on mahdollista, että aihetta on tutkittu tai vaatimusmäärittelyn menetelmiä vastaavia menetelmiä on kehitetty integraatiotilanteen tarpeiden mukaisesti, mutta niistä puhutaan eri käsittein.

Toisen hakukierroksen tuloksissa esiintyi aineistoja, jotka eivät otsikkonsa tai tiivistelmänsä puolesta antaneet olettaa, että ne käsittelisivät tismalleen tutkielman kiinnostuksen kohdetta. Tietojärjestelmien integraatiota käsittelevissä julkaisuissa esiintyi usein termi ”järjestelmän elinkaari” (life-cycle), joka antaa olettaa, että mahdollisiin tuleviin järjestelmäintegraatioihin varautuminen on tärkeä osa tietojärjestelmäkehitysprosessia. Organisaatioiden yhdistyminen voi kuitenkin tapahtua joissakin olosuhteissa yllättäen ja nopealla tahdilla, eikä sen tapahtuessa välttämättä voida olettaa, että integroitavia järjestelmiä kehitettäessä integraatioon olisi varauduttu. Silloin integraatioprosessin säätelyn merkitys korostuu ja erityisesti siirtymävaiheen vaatimusmäärittelylle on tarvetta.

Huomattava osa kolmannen hakukierroksen tuottamista tuloksista käsitteli yrityskaupan jälkeisen tietojärjestelmäintegraation strategiaa (post-merger strategy), johon jatkoselvityksessä kannattaa perehtyä tarkemmin. Jatkokysymyksen K4 mukaisesti lähestymisnäkökulmaa pohtiessa saattaisi olla kannattavaa perehtyä näihin strategioihin ja etsiä viittauksia vaatimusmäärittelyn merkitykseen. Jos jatkotutkimuksesta käy ilmi, että vaatimusmäärittelyllä ei ole aiemmin ollut merkittävää roolia organisaatioiden yhdistymisestä seuraavan integraation strategioissa, olisi varmasti hyödyllistä tutkia erilaisten vaatimusmäärittelyn menetelmien soveltamisen mahdollisuutta integraation eri skenaarioissa ja integraatiostrategian vaiheissa.

## Viiteluettelo

- [Dermeval et al., 2015] Dermeval, D., Vilela, J., Bittencourt, I. I., Castro, J., Isotani, S., Brito, P., & Silva, (2015). Applications of ontologies in requirements engineering: A systematic review of the literature. *Requirements Engineering*, 1-33.
- [Jackson & Stevens, 2000] Jackson, K., & Stevens, R. (2000). Systems Engineering-Models. *Computing & Control Engineering Journal*, 11(4), 187-194.
- [de Lange, 2015] de Lange, T. C. B. (2015). IS/IT integration in mergers & acquisitions: A framework for a systematic approach, Master's thesis, University of Twente.
- [Sommerville, 2007] Sommerville, I. (2007). *Software Engineering* (8th ed.). Harlow: Addison-Wesley.
- [Sommerville & Sawyer, 1997] Sommerville, I., & Sawyer, P. (1997). *Requirements Engineering: A Good Practice Guide* (1st ed.). New York, NY, USA: John Wiley & Sons.

# Sosiaalisten yhteisöjen etsiminen klusterointimenetelmillä

Ville Simpanen

## Tiivistelmä

Sosiaalisiin verkostoihin sisältyvä tietomäärä ihmisistä ja heidän välisistä suhteistaan voi nykyään olla valtava. Suurista verkostoaineistoista voidaan saada tietämystä yhteisöjen löytämismenetelmillä, joilla verkostoja kuvaavista graafeista haetaan solmujen erilaisia ryhmittymiä, joita kutsutaan yhteisöiksi. Yhteisöt samaistetaan graafin klustereihin ja yhteisöjen löytämismenetelmät klusterointimenetelmiin. Tässä tutkielmassa esitellään sellaisia klusterien määritelmiä ja klusterointimenetelmiä, joita sosiaalisten verkostojen tutkimuksissa tavataan yleisimmin.

**Avainsanat ja -sanonnat:** Sosiaalinen verkosto, yhteisöjen löytäminen, klusterointi, tiedon louhinta, tietämyksen muodostaminen tietokannoista.

## 1. Johdanto

Sosiaalisen yhteisön käsite oletetaan intuitiivisesti selväksi. Käsitteellä viitataan esimerkiksi kaupallisiin ja poliittisiin organisaatioihin, perheisiin, ystävyyspiireihin, kyliin, kaupunkeihin, valtioihin, maanosiin ja virtuaalisiin yhteisöihin kuten Facebook-ryhmiin. [Porter *et al.*, 2009]. Toisaalta yhteisön tarkempi määritteleminen on ollut ongelmallisempaa. Tätä kuvaa Hilleryn [1955] vanha katsaus, jossa tutkija käytyään läpi 94 eri määritelmää yhteisölle joutui toteamaan, että ainoa yhteinen piirre näille oli käsitys ihmisistä. Toisaalta lähes kaikki määritelmät sisälsivät jonkinlaisen käsityksen ihmisten välisestä sosiaalisesta kanssakäymisestä.

Hilleryn katsauksen jälkeen ihmisiä ja heidän välisiä suhteita alettiin esittää graafiteorian käsitteillä; alettiin puhua sosiaalisista verkostoista [Scott, 2000]. Sosiaalisten verkostojen tutkimus alkoi myös yleistyä fyysikkojen parissa lähes 50 vuotta katsauksen jälkeen [Porter *et al.*, 2009]. Tähän aikaan Girvan ja Newman [2002] muotoilivat intuition yhteisön sosiaalisista suhteista abstraktimmin osana graafeja. He esittivät, että yhteisöt ovat solmujen ryhmittymiä, joiden sisällä solmujen välisiä särmiä on tiheässä, mutta joiden välillä särmiä on harvassa. Samalla he julkaisivat myös algoritmin tällaisten yhteisöjen löytämiseksi.

Girvanin ja Newmanin tutkimuksen jälkeen erilaiset graafeihin pohjautuvat yhteisöjen löytämismenetelmät ovat olleet kiivaan tutkimuksen kohteena [Porter *et al.*, 2009]. Samalla erilaisia yhteisön määritelmiä on tullut entistä enemmän. Yhteisö ei enää ole vain sosiaalinen käsite, vaan esimerkiksi solujen osien voidaan

nähdä muodostavan yhteisöjä [Fortunato, 2010].

Sosiaalisista verkostoaineistoista löydetty yhteisöt voivat edustaa ihmisryhmiä, joiden jäsenillä on yhteinen mielenkiinto tai jotka ovat esimerkiksi ystäviä keskenään [Girvan and Newman, 2002]. Ne voivat edustaa myös monia muita merkityksellisiä ”sosiaalisia yhteisöjä” [Porter *et al.*, 2009]. Yhteisöjen löytämisellä onkin tärkeä tehtävä sosiaalisten verkostojen analysoimisessa. Verkostoaineistoa on nykyään mahdollista saada valtavia määriä sosiaalisesta mediasta, joka voi sisältää tietoa miljardeista ihmisistä ja heidän välisistä vuorovaikutuksistaan [Hansen *et al.*, 2011]. Jotta tällaisista tietomääristä voitaisiin saada hyödyllistä tietämystä, ovat niiden analysoimisessa avainasemassa juuri tietotekniikkaa hyödynnävät menetelmät [Fayyad *et al.*, 1996].

Tässä tutkielmassa tarkastellaan erilaisia tapoja löytää yhteisöjä klusterointimenetelmillä graafeista, jotka kuvaavat sosiaalisia verkostoja. Luvussa 2 esitellään sosiaalinen verkosto empiiristen tutkimusmenetelmien käsitteenä. Luvussa 3 käydään läpi graafeihin liittyviä peruskäsitteitä ja määritelmiä sekä tarkastellaan, kuinka näillä voidaan merkitä sosiaalista verkostoa. Luvussa 4 määritellään yhteisö klusterina graafissa ja tapana kuvata verkostoaineistoa. Luvussa 5 esitellään sosiaalisten verkostojen tutkimuksissa yleisimmin tavattuja menetelmiä klustereiden löytämiseksi. Lopuksi luvussa 6 on yhteenveto.

## 2. Sosiaalinen verkosto

*Sosiaalinen verkostoanalyysi eli verkostoanalyysi* (social network analysis) käsittää joukon menetelmiä *verkostoaineiston* (social network data) tutkimusta varten. Tällainen aineisto on mitattu *havaintoyksiköistä* (unit of observation), joiden suhteita verkostoanalyysissä pyritään tarkastelemaan. Havaintoyksikkönä on useimmiten jokin henkilö, mutta se voi olla myös jokin muu asia. Verkostoaineisto koostuu ainakin yhdestä *suhdemuuttujasta* (structural variable), joka mittaa jotain yhteyttä eli *suhdetta* (relation) kahden havaintoyksikön välillä. Suhdemuuttujien lisäksi verkostoaineistoon voi sisältyä myös *ominaisuusmuuttujia* (composition variable, actor attribute variable), jotka mittaavat jotain yksittäiseen havaintoyksikköön liittyvää piirrettä. *Sosiaalisesta verkostosta* (social network) voidaan puhua, kun viitataan niihin havaintoyksiköihin ja suhteisiin, joista verkostoaineisto on mitattu. Samalla viitataan myös niihin tapoihin, joilla tällaista aineistoa käsitellään. [Johanson *et al.*, 1995].

Verkostoaineistot voidaan jakaa niiden *ulottuvuuksien* (mode) mukaan. Tämä kertoo, kuinka moneen eri tyyppiin verkostoaineistoon kuuluvat havaintoyksiköt jaetaan. Yksinkertaisimmillaan verkosto on *yksiulotteinen* (unimodal, one-mode), jolloin se koostuu vain yhden toimijajoukon välisistä suhteista. *Toimijat* (agent, actor) yksiulotteisissa verkostoissa ovat tyypillisesti ihmisiä, alueita, yrityksiä tai muita ryhmittymiä. Näihin liittyvät ominaisuusmuuttujat voivat ilmaista piirtei-

tä, kuten ikää, sukupuolta tai yrityksen henkilöstömäärää. [Johanson *et al.*, 1995] Toimijoiden välisillä suhteilla on sisältö, jonka perusteella suhteet voidaan jakaa tyyppeihin, kuten taulukossa 1 on esitelty. Taulukoiduista sisällöistä keskeisenä verkostanalyysin mittauksen kohteena ovat olleet ystävyyssuhteet yksilöiden välillä. Vuorovaikutus suhdetyyppinä voi käsittää myös monia muita eri sisältöjä taulukossa esitettyjen esimerkkien lisäksi. Vuorovaikutusta voi olla myös istuminen jonkun vieressä, osallistuminen samoille kutsuille tai vierailu henkilön kotona jne. [Knoke and Kuklinski, 1982; Johanson *et al.*, 1995] Taulukko 1 kuvaa sisältöjä, joita verkostanalyysin tutkimuksissa yleisimmin operationalisoidaan, eli joista luodaan suhdemuuttujan mittari, joka on sekä luotettava että validi. Knoke ja Kuklinskin mukaan tällaisten sisältöjen lukumäärä on lähes rajaton.

Suhteen tyyppi	Esimerkki sisällöstä
Yksilöllinen arviointi	Ystävyys, pitäminen, kunnioitus
Materiaalisten voimavarojen siirto	Lainaaminen, ostaminen/myyminen
Ei-materiaalisten voimavarojen siirto	Kommunikaatio, auttaminen, informaation lähettäminen ja vastaanottaminen
Vuorovaikutus	Keskustelu, halaaminen, lyöminen
Liikkuminen	Fyysinen liikkuminen (siirtolaisuus ja muuttaminen), sosiaalinen liikkuvuus (muutokset ammattiasemassa ja statuksessa)
Viralliset roolit	Opettajan ja oppilaan suhde, esimiehen ja alaisen suhde
Sukulaisuus	Avoliitto, sukulaisten välinen kanssakäyminen, suvullinen polveutuminen

Taulukko 1: Erilaisia suhdetyyppejä [Johanson *et al.*, 1995].

*Kaksiulotteisessa* (bimodal, two-mode) verkostoaineistossa havaintoyksiköt on jaettu kahteen eri ryhmään. Tällaisessa aineistossa suhteet on mitattu vain eri ryhmiin kuuluvien havaintoyksiköiden välillä. Nämä ryhmät voivat yhtäältä olla molemmat samanlaisia toimijoita kuin yksiulotteisissa verkostoissa. Esimerkiksi toisena joukkona voivat olla liikeyritykset ja toisena vapaaehtoisjärjestöt, jolloin suhde näiden välillä voi kuvata sitä, miten liikeyritykset antavat järjestöille lahjoituksia. Toisaalta kaksiulotteinen verkosto voi koostua vain yhdestä toimijajoukosta ja näiden lisäksi tapahtumista, joihin toimijat osallistuvat. Tällaista verkostoa kutsutaan myös *jäsenyyssverkostoksi* (affiliation network). Jäsenyyssverkoston tapahtuma voidaan määritellä laajasti, ja se voi kattaa eri asioita aina epämuodollisista kokoontumisista virallisiin organisaatiojäsenyyksiin. Toimijan suhde tapahtumaan voi merkitä esimerkiksi yksilön kuulumista johonkin yhdistykseen tai jonkin maan kuulumista kansainväliseen järjestöön. [Johanson *et al.*, 1995] Hansenin ja muiden [2011] mukaan toimijat voivat olla jäsenyyssverkostossa liittyneinä myös johonkin aktiviteettiin tai muuhun sisältöön. *Yhteistyöverkosto* (collaboration network) on jäsenyyssverkosto, jossa kahden toimijan suhde sisältöön kuvaa heidän

välistä yhteistyötä sisällön tuottamisessa [Newman, 2003]. Verkostoaineisto voi olla myös *moniulotteinen* (multimodal), jos sen havaintoyksiköt on jaettu useampaan kuin kahteen tyyppiin. Moniulotteisia verkostoaineistoja on kuitenkin tutkittu vain harvoin niiden käsittelemisen hankaluuden takia. [Johanson *et al.*, 1995]

Verkostoaineistoa voidaan kerätä eri tavoin. Perinteisiä menetelmiä sosiaalisten verkostojen tutkimuksessa ovat olleet esimerkiksi haastattelut ja kyselylomakkeet. Näissä tietoa sosiaalisista suhteista on voitu kysyä suoraan tutkimukseen osallistuvilta henkilöiltä. [Scott, 2000] Tämä kuitenkin aiheuttaa tutkimuksen kannalta useita epävarmuustekijöitä. Scottin mukaan sosiaaliset suhteet määräytyvät tilannekohtaisesti, eikä tutkija siten voi olla varma, että jokaisella haastateltavalla on ollut samanlainen käsitys tutkimuksessa haetun suhteen kuten ystävyyden merkityksestä. Toisaalta suhteen merkitys voi jäädä keinotekoiseksi, jos tutkija itse asettaa sille jonkin kriteerin, esimerkiksi tapaamisien toistumistiheyden. Perinteisiä verkostoaineiston keräystapoja on myös muita, kuten havainnoiminen. [Scott, 2000] Näiden tapojen yhteinen piirre on kuitenkin niiden manuaalisuus, joka rajoittaa tutkittavan verkoston kokoa [Newman, 2003].

Perinteisen tiedonkeruun rajoitteet ovat ajaneet tutkijoita hankkimaan tietoa sosiaalisista verkostoista muilla tavoin. Eräs jokseenkin luotettava tapa kerätä verkostoaineistoa on käyttää tietoa ihmisten välillä tapahtuneesta viestinnästä. [Newman, 2003] Tällaista tietoa voidaan saada esimerkiksi *sosiaalisesta mediasta* (social media), jolla viitataan yleisesti erilaisiin verkkosovelluksiin, jotka tukevat ihmisten välistä vuorovaikutusta [Hansen *et al.*, 2011]. Sosiaalisen median sovellus tallentaa tyypillisesti joukon tapahtumia (transaktioita), joihin liittyy eri havaintoyksiköitä. Näin suhteen katsotaan muodostuvan niiden havaintoyksiköiden välille, jotka liittyvät samaan tapahtumaan. [Papadopoulos *et al.*, 2012] Perinteinen esimerkki tällaisesta joukosta tapahtumia on loki sähköpostiviestinnästä. Kahden henkilön välillä voidaan katsoa olevan suhde, jos he ovat lähettäneet viestin toisilleen. Sähköpostiviestinnän perusteella luodun verkoston voidaan katsoa kuvaavan tarkasti sen toimijoiden suhteita reaali maailmassa sähköpostin yleisyyden takia. Varsinkin nytemmin yleistynyt tapa on kerätä tietoa viestinnästä myös erilaisista *yhteisöpalveluista* (social networking site), kuten Facebookista tai Twitteristä. [Hansen *et al.*, 2011; Fortunato, 2010].

Oleellinen piirre sosiaalisia verkostoja käyttävissä tutkimuksissa on se, että kiinnostuksen kohteina ovat juuri suhteet eivätkä niinkään havaintoyksiköihin liittyvät ominaisuudet. [Scott, 2000] Sosiaalisen verkoston käsite on laajentunut kattamaan myös muiden tieteenalojen tarpeita, ja erilaisia ”verkostoja” on alettu käyttää esimerkiksi fysikaalisten ilmiöiden tutkimuksissa [Papadopoulos *et al.*, 2012]. Newman [2003] jakaa verkostot löyhästi neljään eri kategoriaan: sosiaaliin, teknologisiin ja biologisiin verkostoihin sekä informaatioverkostoihin. Nämä verkostot säilyttävät intuition siitä, että verkosto koostuu havaintoyksiköistä ja näiden

välisistä suhteista, mutta eri tieteenalat määrittelevät näille omia tarkempia merkityksiään [Porter *et al.*, 2009]. Esimerkiksi biologisissa verkostoissa havaintoyksiköt voivat olla solujen osia ja suhteet voivat kuvata näiden osien toimintaa samoissa soluissa [Fortunato, 2010].

### 3. Verkostojen kuvaaminen graafina

Verkostoaineistoja pyritään kuvaamaan merkintätavoilla, jotka mahdollistavat näitä analysoivien menetelmien tarkan kuvauksen. Eräänä formaalina merkintätapana voidaan käyttää graafeja [Johanson *et al.*, 1995]. Graafit ovat matemaattisia rakenteita, joita tutkivaa matematiikan haaraa kutsutaan graafiteoriaksi. Itse verkosto voidaan samaistaa graafiin, jolloin termiä käytetään abstraktina yläkäsitteenä viittaamaan eri tyyppisiin verkostoihin [Porter *et al.*, 2009]. Graafeja on eri tyyppisiä, joista jokainen asettaa kuvattavalle aineistolle erilaisia vaatimuksia. Tässä käydään läpi sellaisia graafityyppejä, joita sosiaalisten verkostojen tutkimuksessa yleisimmin tavataan. Monesti näihin viitataan kirjallisuudessa eri termein [Koivisto ja Niemistö, 2001]. Tässä määritellyt termit ja peruskäsitteet pohjautuvat pääasiassa Vasudevin [2007] kirjaan, ja näiden suomennoksissa on käytetty Koiviston ja Niemistön [2001] monistetta.

#### 3.1. Eri tyyppisiä graafeja

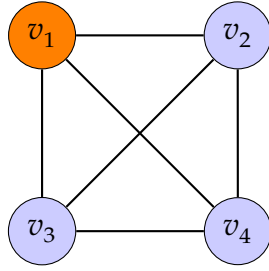
*Graafi* (graph) koostuu joukosta  $V$ , jonka alkioita kutsutaan solmuiksi, ja joukosta  $E$ , jonka alkiot ovat särmiä. Vastaavasti joukkoa  $V$  kutsutaan solmujoukoksi ja joukkoa  $E$  särmäjoukoksi. Graafia merkitään usein järjestettynä parina  $G = (V, E)$ .

*Suuntaamaton graafi* (undirected graph, graph)  $G = (V, E)$  on graafi, jossa jokaiseen särmään  $e \in E$  liittyy järjestämätön solmupari  $\{u, v\}$ , jossa  $u, v \in V$ . Jos nämä parit olisivat järjestettyjä, olisi kyseessä *suunnattu graafi* (directed graph) eli digraafi. Suunnatun graafin tapauksessa särmiä kutsutaan myös kaariksi. Sekä suunnatuille että suuntaamattomille graafeille määritellään lisäksi, että

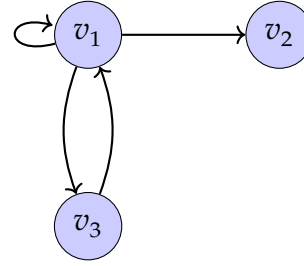
- solmu  $u$  on solmun  $v$  *vierussolmu* (adjacent vertex, neighbour),
- solmut  $u$  ja  $v$  ovat särmän  $e$  *päätesolmut* (end vertices, end points),
- särmä  $e$  *kulkee* (incident with) solmujen  $u$  ja  $v$  kautta ja samoin
- särmä  $e$  *yhdistää* (connects) solmut  $u$  ja  $v$ .

Suunnattuja ja suuntaamattomia graafeja voidaan visualisoida pisteillä ja näiden välisillä viivoilla siten, että pisteet vastaavat graafin solmuja ja viivat särmiä. [Vasudev, 2007] Kuvassa 1 on esimerkkejä erilaisista graafeista, joita tässä luvussa käsitellään.

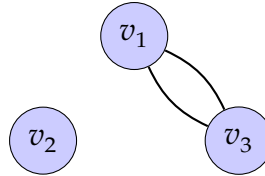
(a) Yksinkertainen graafi, klikki.



(b) Suunnattu pseudograafi.



(c) Suuntaamaton multigraafi.



Kuva 1: Erilaisia graafeja.

Kun sosiaalista verkostoa kuvataan graafilla, valitaan solmuiksi tyypillisesti sen havaintoyksiköt ja särmiksi näiden väliset suhteet. Suhteisiin liittyy ominaisuuksia, jotka määräävät, minkä tyyppisellä graafilla niitä voidaan kuvata ilman, että menetetään tietoa. Tällainen ominaisuus on esimerkiksi suunta, jonka perusteella verkostoa voidaan merkitä joko suuntaamattomalla tai suunnatulla graafilla. Suuntaamattomat suhteet eivät erottele, kumpi suhteeseen liittyvistä toiminnoista on suhteessa kumman kanssa. Esimerkiksi, jos Matti ja Ville käyvät samaa koulua, niin ei todennäköisesti tarvitse erottaa, käykö Ville Mattin koulussa vai Matti Villen koulussa. [Scott, 2000] Toisaalta esimerkiksi vain vanhemmat voivat olla lastensa äitejä tai isiä eikä päinvastoin, jolloin suhde on suunnattu [Johanson *et al.*, 1995].

Suhteeseen voi liittyä jokin määre, jolla pyritään kuvaamaan sen frekvenssiä, voimakkuutta tai muunlaista luonnetta. Esimerkiksi yksilöllinen arviointi (ks. taulukko 1) ilmaisee monesti toimijan joko negatiivista tai positiivista tuntemusta jostain henkilöstä tai asiasta kohtaan. [Johanson *et al.*, 1995] Tämä voidaan esittää graafin särmiin liitettyinä plus- tai miinusmerkkeinä. Toisaalta suhteella voi olla jokin numeerinen arvo, joka tavalla tai toisella kuvaa suhteen voimakkuutta. Arvo voi edustaa esimerkiksi sitä, kuinka monta kertaa kaksi henkilöä ovat tavanneet viikon aikana. Tällainen arvo voidaan esittää *painotetussa graafissa* (weighted graph), jossa jokaiseen särmään on liitetty reaaliluku eli *paino* (weight) [Scott, 2000; Vasudev, 2007]. Arvotettujen suhteiden vastakohtana ovat dikotomiset suhteet, jotka vain joko ovat olemassa tai eivät.

Kaksiulotteisen verkoston ominaispiirteet voidaan merkitä *kaksijakoisella* (bipartite) graafilla. Graafi  $(V, E)$  on kaksijakoinen graafi, jos solmujoukko  $V$  voidaan jakaa kahteen sellaiseen epätyhjään joukkoon  $V_1$  ja  $V_2$ , että jokaisen särmän



päätesolmuista toinen kuuluu joukkoon  $V_1$  ja toinen joukkoon  $V_2$ . [Koivisto ja Niemistö, 2001; Vasudev, 2007] Samoin moniulotteista verkostoa vastaa  $k$ -jakoinen ( $k$ -partite, multipartite) graafi. Suuntaamaton graafi  $(V, E)$  on  $k$ -jakoinen, jos on olemassa sellainen joukon  $V$  ositus  $\{V_1, V_2, \dots, V_k\}$ , etteivät joukon  $E$  minkään särmän molemmat päätesolmut kuulu samaan joukkoon  $V_i$  ( $1 \leq i \leq k$ ). [Koivisto ja Niemistö, 2001]

Seuraavat määritelmät edustavat särmien erikoistapauksia:

- Jos särmän molemmat päätesolmut ovat samat, se on *luuppi* (loop).
- Suuntaamattoman graafin kaksi särmää ovat *rinnakkaiset* (parallel edges, multiple edges), jos ne kulkevat kahden saman solmun kautta. Suunnatussa graafissa särmiin liittyvien järjestettyjen parien täytyy olla samat.

Sellaista graafia, joka sallii sekä rinnakkaissärmät että luupit kutsutaan myös *pseudograafiksi* (pseudo graph). Kaikki sosiaaliset suhteet eivät anna luupeille mielekästä tulkintaa [Scott, 2000]. Esimerkiksi suhde ”käy samaa koulua” triviaalisti pätee, jos sen molemmat osapuolet ovat samat. Jos luuppeja ei tarvitse merkitä, voidaan verkostoa käsitellä *multigraafina* (multigraph). Multigraafi on pseudograafi (eli graafi), joka ei siis salli luuppeja [Vasudev, 2007]. Nimi kuvastaa sitä, että se sallii rinnakkaiset särmät. Näillä voidaan kuvata esimerkiksi eri sisältöjä, joita sosiaalisilla suhteilla voi olla [Hansen *et al.*, 2011].

Se, minkälaista graafia tutkimuksessa päädytään tarkastelemaan, riippuu käytettävän aineiston lisäksi siitä, mitkä ovat tutkimuksen vaatimukset ja mitä verkoston analyysimenetelmiä on käytettävissä. Verkostoja voidaan yksinkertaistaa, jotta useampia menetelmiä voitaisiin soveltaa, vaikkakin tässä menetetään tietoa niiden rakenteesta [Scott, 2000; Papadopoulos *et al.*, 2012]. Esimerkiksi sen sijaan, että jokaista suhdetyyppiä merkittäisiin rinnakkaissärmällä, voidaan suhteista ottaa tarkasteltavaksi vain osa. Toisaalta rinnakkaissärmät voidaan myös yhdistää tavalla tai toisella yhdeksi (painotetuksi) särmäksi. [Fortunato, 2010]. Jos taas graafi sisältää painotettuja särmiä, voidaan näitä muuttaa painottomiksi. Särmiä voidaan esimerkiksi ottaa tai olla ottamatta painottamattomaan graafin jonkin painojen kynnysarvon perusteella. Vastaavasti suhteiden suunta voidaan jättää merkitsemättä, jos merkittävää suhteessa on vain se, onko se olemassa. [Scott, 2000]

Jäsenyysverkosto voidaan yksinkertaistaa yksiulotteiseksi verkostoksi, jossa toimijoiden välillä on suhde aina, kun ne liittyvät jäsenyysverkostossa samaan havaintoyksikköön. Yhteistyöverkостossa havaintoyksiköitä voivat olla esimerkiksi wikisivut ja toimijoita wikipalvelun käyttäjät. Suhde käyttäjän ja sivun välillä ilmaisee, että käyttäjä on muokannut sivua. Tällöin vastaavassa yksiulotteisessa verkostossa suhde esiintyy käyttäjien välillä, jos he ovat muokanneet samaa sivua. [Hansen *et al.*, 2011]

### 3.2. Muita graafeihin liittyviä käsitteitä

Solmun *aste* (degree) kuvaa, kuinka monen särmän päätesolmuna solmu on. Jos solmun kautta kulkee luuppi, sen aste on ainakin 2. [Koivisto ja Niemistö, 2001] Jos graafin  $G$  solmujoukko on  $V = \{v_1, \dots, v_n\}$ , niin graafin  $G$  *astejono* (degree sequence) on jono  $k_1, \dots, k_n$ , missä  $k_i$  ( $1 \leq i \leq n$ ) tarkoittaa solmun  $v_i$  astetta. [Fortunato, 2010]

Graafin solmujoukon  $V$  ositusta kahteen osajoukkoon  $S$  ja  $V/S$ , kutsutaan solmujoukon  $V$  *irrotukseksi* (cut). Tämän *irrotuksen koko* (cut size) on niiden särmien lukumäärä, joiden toinen päätesolmu kuuluu joukkoon  $S$  ja toinen joukkoon  $V/S$ . Jos särmät ovat painotettuja, voidaan irrotuksen koko määritellä myös näiden painojen summana. [Schaeffer, 2007]

Solmun  $v$  kaikkien vierussolmujen joukko on solmun  $v$  *naapurusto* (neighbourhood), jota merkitään  $\Gamma(v)$ . Vastaavasti solmujoukon osajoukon  $U$  naapurusto on tämän solmujen kaikkien naapurustojen yhdisteen joukko  $\{\Gamma(u) \mid u \in U\}$ . [Fortunato, 2010]

*Polku* (path, walk) suuntaamattoman graafin solmusta  $v_0$  solmuun  $v_k$  on äärellinen jono

$$v_0, e_1, v_1, e_2, v_2, \dots, e_k, v_k,$$

jossa jokaista peräkkäistä solmuparia  $v_{i-1}, v_i$  ( $1 \leq i \leq k$ ) yhdistää särmä  $e_i$ . Suunnatussa graafissa jokainen särmä  $e_i$  on järjestetty solmupari  $(v_{i-1}, v_i)$ , eli polkua voi kulkea vain yhteen suuntaan. Kaksi graafin solmua ovat *yhdistetyt* (connected), jos niiden välillä on polku.

Suuntaamaton graafi on *yhtenäinen* (connected), jos sen minkä tahansa kahden solmun välillä on polku. Suunnatussa graafissa yhtenäisyys määritellään eri tavoin ja se jaetaan heikkoon ja vahvaan. Olkoon graafi  $G$  suuntaamaton graafi, joka saadaan muuntamalla kaikki suunnatun graafin  $H$  särmät suuntaamattomiksi. Suunnattu graafi  $H$  on heikosti yhtenäinen, jos graafi  $G$  on yhtenäinen. Suunnattu graafi on vahvasti yhtenäinen, jos graafin jokaisen kahden solmun  $u$  ja  $v$  välillä on polku sekä solmusta  $u$  solmuun  $v$  että solmusta  $v$  solmuun  $u$ . [Koivisto ja Niemistö, 2001]

Graafi  $H = (W, F)$  on graafin  $G = (V, E)$  *aligraafi* (subgraph), jos pätee  $W \subseteq V$  ja  $F \subseteq E$ . Aligraafin  $H$  kaikki särmät ja solmut kuuluvat siten graafiin  $G$ . Jos lisäksi  $W \subset V$  tai  $F \subset E$  niin kyseessä on *aito aligraafi* (proper subgraph).

Graafin  $G$  aligraafia  $H$  sanotaan graafin  $G$  *maksimaaliseksi* (maximal) aligraafiksi jonkin ominaisuuden  $P$  suhteen, jos

- (i) graafilla  $H$  on ominaisuus  $P$  ja
- (ii) aina, kun  $H$  on graafin  $G$  mielivaltaisen aligraafin  $F$  aito aligraafi, aligraafilla  $F$  ei ole ominaisuutta  $P$ . [Koivisto ja Niemistö, 2001]

Aligraafi on siten jonkin ominaisuuden suhteen maksimaalinen, jos siihen ei voida lisätä alkuperäisestä graafista yhtään solmua tai särmää ilman, että tämä ominaisuus häviää. Graafin maksimaalista yhtenäistä aligraafia kutsutaan graafin *komponentiksi* (component) [Koivisto ja Niemistö, 2001; Vasudev, 2007].

Olkoon  $G = (V, E)$  graafi ja  $W$  sen epätyhjä solmujoukko. Tällöin  $H = (W, F)$  on joukon  $W$  (*solmu*)*indusoima* (vertex-induced) graafin  $G$  aligraafi, jos  $F \subseteq E$  muodostuu niistä joukon  $E$  särmistä, joiden päätesolmut kuuluvat joukkoon  $W$ . Tällainen aligraafi on yksikäsitteinen ja se samaistetaan joskus solmujoukkoonsa.

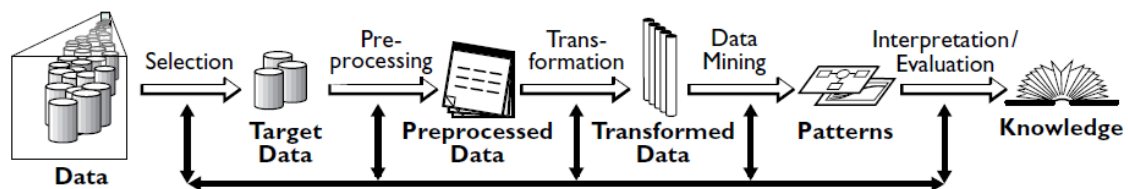
Graafia  $G = (V, E)$  jossa ei ole luuppeja eikä rinnakkaissärmiä, kutsutaan yksinkertaiseksi graafiksi. [Vasudev, 2007]. Suuntaamaton yksinkertainen graafi on *täydellinen* (complete) graafi, jos sen minkä tahansa kahden solmun välillä kulkee särmä. Jos aligraafi on täydellinen, niin sitä kutsutaan myös *klikiksi* (clique). [Koivisto ja Niemistö, 2001] Verkostanalyysissä on lisäksi yleistä vaatia, että tällainen aligraafi on maksimaalinen. [Fortunato, 2010] Maksimaalisen ja ei-maksimaalisen klikin välistä eroa on havainnollistettu kuvassa 1a. Olkoon  $G$  kuvassa esitetty graafi. Graafi  $G$  on klikki ja tällä ominaisuudella itsensä maksimaalinen aligraafi. Lisäksi käytetään solmujen  $\{v_2, v_3, v_4\}$  indusoimasta graafin  $G$  aligraafista merkinää  $H$ . Myös aligraafi  $H$  on klikki, mutta se ei ole maksimaalinen tällä ominaisuudella graafissa  $G$ .

#### 4. Yhteisöt graafeissa

Tutkimuksessa kerätyn aineiston arvokkuus ilmenee pääasiassa vasta siinä, kun aineistosta voidaan havaita mielenkiintoisia tapahtumia tai trendejä, kun siitä saadaan apua päätöksentekoa varten tai kun aineistoa voidaan käyttää hyväksi esimerkiksi jotakin kaupallista tai tieteellistä määränpäättä varten. Toisin sanoen aineiston arvokkuus näkyy vasta sitten, kun siitä saadaan *tietämystä* (knowledge). Perinteisesti tietämystä on pyritty muodostamaan manuaalisesti. Manuaalisten menetelmien ongelmana on kuitenkin niiden hitaus, kalleus ja hyvin subjektiivinen luonne. Sosiaalinen media on mahdollistanut hyvin suurien sosiaalisten verkostojen luomisen, jotka voivat koostua miljoonista eri havaintoyksiköistä [Fortunato, 2010]. Näin valtavia aineistoja ei käytännössä edes pystytä analysoimaan manuaalisilla menetelmillä. [Fayyad *et al.*, 1996]

Manuaalisille aineistojen analysointimenetelmille on pyritty löytämään vaihtoehtoja tietotekniikkaa hyödyntävistä ratkaisuista. Erään vaihtoehdon kuvaa *tietämyksen muodostaminen tietokannoista* (knowledge discovery in databases, KDD). Tämä on prosessi (tästä eteenpäin myös KDD-prosessi), joka kattaa askeleet mahdollisesti suurten tietomäärien hankinnasta siihen, kuinka aineistosta voidaan saada lopulta tietämystä. Fayyad ja muut [1996] kuvaavat KDD-prosessia ”epätriviaalina prosessina, jossa pyritään muodostamaan valideja, uusia, potentiaalisesti käyttökelpoisia ja lopulta ymmärrettäviä hahmoja aineistosta”. Tässä hahmolla

tarkoitetaan jotakin tapaa kuvata tai mallintaa aineistoa. Löydettyjen hahmojen tulisi olla valideja, eli niiden tulisi kuvata sitä, mitä tutkimuksessa haluttiinkin kuvata. Prosessi viittaa siihen, että tietämyksen muodostaminen koostuu useasta eri askeleesta. KDD-prosessi pyrkii vähentämään aineistojen tulkintaan liittyviä ongelmia. Sekään ei kuitenkaan täysin vältty subjektiivisilta piirteiltä, sillä ominaisuudet kuten validius, käyttökelpoisuus ja ymmärrettävyys eivät ole yksiselitteisiä. [Fayyad *et al.*, 1996]



Kuva 2: Yleiskuva tietämyksen muodostamisen vaiheista [Fayyad *et al.*, 1996].

KDD-prosessin askelia on havainnollistettu kuvassa 2. Ensimmäiset askeleet ennen *tiedonlouhintaa* (data mining) vastaavat esimerkiksi verkostoaineiston keräystä, tämän esitysmuodon valintaa ja mahdollista yksinkertaistamista. Lisäksi askeliin voi kuulua myös muita aineiston esikäsittelytapoja, kuten virheellisten mittaustulosten vähentämistä, mutta näitä tapoja ei tässä tutkielmassa käsitellä. Tiedonlouhinnan askel on se vaihe, jossa hahmoja pyritään löytämään aineistosta. Tämän jälkeen löydettyjä hahmoja pyritään tulkitsemaan ja osana tätä mahdollisesti visualisoimaan. Prosessi ei ole täysin suoraviivainen, vaan sen aikana voidaan palata aikaisempiin askeliin, mitä on kuvattu mustilla nuolilla. [Fayyad *et al.*, 1996].

Paljon viime aikoina tutkittu sosiaalisten verkostojen tiedonlouhintatehtävä on ollut *yhteisöjen löytäminen* (community detection) eli *graafien klusterointi* (graph clustering), jossa tavoitteena on ryhmitellä graafin solmuja tyypillisesti siten, että ryhmien sisällä särmiä esiintyy ”tiheässä” ja välillä ”harvassa”. Tällaiset ryhmät voivat sosiaalisten verkostojen tapauksessa edustaa esimerkiksi sellaisia käsitteitä kuten perhettä, ystävyyspiirejä, kaupunkeja tai valtioita. Solmujen ryhmittely perustuu pääasiassa särmiin liittyvään eli *rakenteelliseen* (structural, topological) tietoon. Toisaalta solmuja voidaan ryhmitellä myös niiden sisäisen tiedon pohjalta. Tällöin ryhmien intuitio on ennemminkin se, että niiden sisällä olevien solmujen tulisi olla jollakin tapaa samankaltaisia [Jain *et al.*, 1999]. Graafien klusteroinnissa saatuja solmujen ryhmittymiä eli hahmoja kutsutaan *yhteisöiksi* (community) tai *klustereiksi* (cluster). [Fortunato, 2010]

Yleisesti määriteltynä klusteri graafissa  $G$  on tämän solmujoukon tietyn osajoukon indusoima graafin  $G$  aligraafi. Klusteri sisältää siten vain ja ainoastaan ne

särmät, jotka yhdistävät sen solmuja graafissa  $G$ . Usein klusterina pidetään myös itse solmujoukkoa. Lisäksi voidaan vaatia, että klusterit ovat yhtenäisiä. Klusteri voidaan tarkemmin määritellä useilla eri tavoilla, joiden tyyppejä esitellään seuraavaksi. Yleensä klusterit eivät kuitenkaan perustu mihinkään erilliseen määritelmään, vaan klustereiksi vain määritellään ne aligraafit, jotka jokin menetelmä tuottaa [Fortunato, 2010; Schaeffer, 2007].

Eräs tapa määritellä klusteri on käyttää jotakin *läheisyysmittaa* (proximity measure, similarity measure). Läheisyysmitta määrittelee graafissa klusterit, joissa kaikki solmut ovat mahdollisimman lähellä muita klusterin solmuja. Läheisyyden yhteydessä voidaan puhua myös *samankaltaisuudesta* (similarity) tai *etäisyydestä* (distance), jolloin samaan klusteriin sopivat paremmin suuremman samankaltaisuuden ja vastaavasti pienemmän etäisyyden omaavat solmut. Läheisyysmitat voivat pohjautua perinteiseen *tiedon klusterointiin* (data clustering), jolloin mitan arvo lasketaan klusteroitaviin solmuihin liitettyjen  $n$ -ulotteisten vektoreiden pohjalta [Jain *et al.*, 1999; Fortunato, 2010]. Vektorin alkiot vastaavat tällöin solmuun liitettyjä ominaisuusmuuttujia.

Tiedon klusteroinnissa läheisyysmittoja on kehitetty monia, sillä niiden soveltuvuus riippuu paljolti kohdealueesta [Schaeffer, 2007]. Jos  $\mathbf{d}_i$  ja  $\mathbf{d}_j$  ovat havaintoyksiköihin liitettyjä vektoreita, niin etäisyysmitan riittää toteuttaa seuraavat ehdot:

- Havaintoyksikön etäisyys itseensä on 0:  $\text{dist}(\mathbf{d}_i, \mathbf{d}_i) = 0$ .
- Etäisyys on symmetrinen:  $\text{dist}(\mathbf{d}_i, \mathbf{d}_j) = \text{dist}(\mathbf{d}_j, \mathbf{d}_i)$ .
- Etäisyydelle pätee kolmioepäyhtälö:  $\text{dist}(\mathbf{d}_i, \mathbf{d}_j) \leq \text{dist}(\mathbf{d}_i, \mathbf{d}_k) + \text{dist}(\mathbf{d}_k, \mathbf{d}_j)$ .

Kahden vektorin välinen etäisyys määritellään tyypillisesti *euklidisena etäisyytenä* (euclidean distance), jos näiden alkiot ovat jatkuvia muuttujia [Schaeffer, 2007]. Kun vektorit ovat kaksiulotteisia, tämä vastaa kahden tasolla olevan pisteen etäisyyttä. Euklidinen etäisyys voidaan määritellä yleisemmin  $n$ -ulotteisten vektorien tapauksessa kaavalla

$$\text{dist}(\mathbf{d}_i, \mathbf{d}_j) = \sqrt{\sum_{k=1}^n (d_{i,k} - d_{j,k})^2}.$$

Aina graafin solmuihin ei voida liittää vektoreita läheisyyden laskemista varten, esimerkiksi jos aineistoon ei liity yhtään ominaisuusmuuttujaa. Läheisyysmitta voidaan tällöin perustaa graafin rakenteeseen liittyvään tietoon. Graafien klusteroinnille ominaisempaa onkin johtaa kahden solmun läheisyys niihin liittyvien särmien perusteella. Yksinkertainen tapa on esimerkiksi laskea, kuinka monta yhteistä vierussolmua solmuilla on suhteessa niiden molempien vierussolmujen kokonaismäärään. Tällainen solmujen  $u, v$  samankaltaisuus  $\omega(u, v)$  voidaan laskea kaavalla

$$\omega(u, v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}.$$

Mitta saa arvot väliltä  $[0, 1]$ : arvo on 0, jos solmuilla ei ole yhteistä vierussolmua ja 1 kun vierussolmut ovat samat. [Schaeffer, 2007]

Yhteisöjen määrittelytapoja voidaan jakaa *globaaleihin* (global) ja *lokaaleihin* (local) sen perusteella, käyttävätkö ne tietoa koko graafista vai vain jostakin sen (aidosta) aligraafista. Globaali määritelmä on tyypillistä perustaa *modulaarisuuteen* (modularity), jonka arvo lasketaan kuvaamaan löydettyjen klusterien hyvyttä. Klustereiksi voidaan määritellä tällöin se joukko, joka tuottaa parhaimman arvon. [Fortunato, 2010] Modulaarisuutta ja sen käyttöä klusteroinnissa käsitellään tarkemmin luvussa 5. Myös edellä käsitellyt läheisyysmittoihin perustuvat määritelmät käyttävät tietoa koko graafista.

Yksinkertainen lokaali tapa määritellä klusteri on määritellä se esimerkiksi klikkinä. Tällainen tapa on tunnettu sosiaalisten verkostojen tutkimuksista, joissa klusteri voidaan tulkita esimerkiksi ryhmänä, jonka kaikki jäsenet ovat ystäviä keskenään. Määritelmä on kuitenkin Fortunaton [2010] mukaan liian tiukka, koska tällöin ryhmän ulkopuolelle jäävät kaikki ne, joilta puuttuu yksikin suhde ryhmän muihin jäseniin. Lisäksi suuremmat kuin kolmesta solmusta koostuvat klikit ovat sosiaalisissa verkostoissa harvinaisia. Toinen lokaali tapa määritellä klusteri on vaatia, että graafin sellaisen irrotuksen koko on pieni, jossa klusteri on toisena osajoukkona. [Fortunato, 2010] Irrotuksen käyttöä klusterin määrittelemisessä käsitellään luvussa 5. Lokaalien määritelmien etuna globaaleihin menetelmiin verrattuna on se, että niihin perustuvat menetelmät eivät välttämättä kärsi suorituskykyyn liittyvistä ongelmista verkoston koon kasvaessa [Papadopoulos *et al.*, 2012].

## 5. Graafien klusterointimenetelmät

Jos klustereiden määritelmiä on paljon, niin on myös klustereita löytäviä menetelmiä. Itse asiassa jokaisen määritelmän klustereita pyrkii tyypillisesti löytämään vähintään yksi menetelmä [Papadopoulos *et al.*, 2012]. Kuten klustereiden määritelmät, myös klusterointimenetelmät voidaan jakaa globaaleihin ja lokaaleihin menetelmiin [Schaeffer, 2007]. Tällöin jako perustuu siihen, klusteroivatko ne graafin jokaisen solmun vai näiden tietyn osajoukon. Sosiaalisten verkostojen tutkimuksissa globaalit menetelmät ovat olleet merkittävimpiä, ja tässä luvussa keskitytään näihin.

Klusterointimenetelmien tuottamaa tulosta eli klusterijoukkoa kutsutaan *klusteroinniksi* (clustering). Yhteisön löytämisestä puhuttaessa klusterointiin viitataan myös *yhteisörakenteena* (community structure). Globaalin klusteroinnin tuloksena saadaan solmujen jako klustereihin, joiden solmujoukot muodostavat joko graafin solmujoukon *osituksen* (partition) tai *peiton* (cover). Osituksessa jokainen klusteroitu alkio kuuluu vain yhteen klusteriin, mutta peitossa alkio voi kuulua useampaankin. [Schaeffer, 2007] Peittoon kuuluvien yhteisöjen sanotaan olevan *limittäi-*

*siä* (overlapping). Limittäisiä yhteisöjä hakevat menetelmät ovat saaneet paljon suosiota, sillä yksilön voidaan ajatella kuuluvan useaan ryhmään samaan aikaan, esimerkiksi perheeseen ja ystävyyspiireihin. Näitä tuottavia menetelmiä on kuitenkin tutkittu vähemmän kuin menetelmiä, joiden tuloksena on ositus. [Fortunato, 2010]

Klusteroinnin ja yhteisöjen löytämisen erona voidaan pitää sitä, että yhteisöjen löytäminen menetelmät eivät vaadi tyypillisesti ennakkotietoa yhteisörakenteeseen kuuluvien yhteisöjen lukumäärästä, kun taas klusterointimenetelmät monesti tätä vaativat. Yhteisöjen löytäminen menetelmät sopivat siten paremmin suurien verkostojen analysoimiseen, joissa yhteisöjen määrää on vaikea arvioida etukäteen. [Papadopoulos *et al.*, 2012]

### 5.1. Hierarkkinen klusterointi

Klusterointimenetelmiä, jotka tuottavat yhden osituksen tai peiton, kutsutaan *tasoklusterointimenetelmiksi* (flat clustering method). Klusteroinnin tuloksena voi olla myös useita eri *tasoja* (level), joista jokainen käsittää alkioden jaon (tyypillisesti osituksen). Nämä tasot ovat keskenään hierarkkisessa suhteessa siten, että ylemmälle tasolle kuuluvat klusterit sisältävät yleensä alemman tason klusterit. Tällaisia tasoja tuottavia menetelmiä kutsutaan *hierarkkiseksi klusterointimenetelmiksi* (hierarchical clustering method). [Schaeffer, 2007]

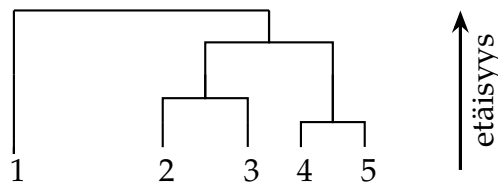
Hierarkkiset klusterointimenetelmät ovat eräitä yleisimmin käytettyjä menetelmiä sosiaalisten verkostojen yhteisöjen löytämiseen [Fortunato and Castellano, 2009]. Tähän ne sopivat esimerkiksi siksi, etteivät ne vaadi tietoa klusterien lukumäärästä. Hierarkkisuus voidaan nähdä myös tyypilliseksi ihmisryhmille. [Fortunato, 2010] Esimerkiksi lapset voitaisiin yhdellä tasolla ryhmitellä heidän käymiensä koulujen perusteella, mutta näitä kouluja vastaavien klusterien sisällä ryhmittelyä voi vielä jatkaa heidän koululuokkiensa mukaan.

Hierarkkiset klusterointimenetelmät voidaan jakaa sen mukaan, tuottavatko ne ylemmän tason yhdistämällä alemman tason klustereita vai alemman tason pilkkomalla klustereita ylemmällä tasolla. *Alhaalta-ylös* (bottom-up) menetelmät eli *kokoavat* (agglomerative) menetelmät yhdistävät klustereita iteratiivisesti suurempiin osiin, kunnes klusterointia ei enää voida jatkaa tai saadut klusterit täyttävät jonkin *pysäytysehdon* (stop condition) eli kriteerin, jolloin klusterointi on saavutettu. Vastaavasti *ylhäältä-alas* (top-down) menetelmät eli *jakavat* (divisive) menetelmät pilkkovat solmujoukkoa iteratiivisesti tai rekursiivisesti. [Schaeffer, 2007]

Kokoavat menetelmät asettavat alkiot klustereihin tyypillisesti jonkin niiden välille määritellyn läheisyysmitan perusteella. Kokoavat menetelmät jaetaan sen mukaan, miten ne soveltavat mittaa klustereihin, jotka sisältävät useamman kuin yhden alkion. *Yhden yhteyden* (single-link) menetelmässä kahden klusterin väliseksi etäisyydeksi määritellään pienin etäisyys, joka esiintyy kahden näissä ole-

van alkion välillä (toinen ensimmäisessä klusterissa ja toinen toisessa). Sitä vastoin *täyden yhteyden* (complete-link) menetelmässä klusterien välinen etäisyys on suurin alkioiden välinen etäisyys. Molemmissa tapauksissa kokoavat klusterointimenetelmät pyrkivät muodostamaan uuden klusterin aina yhdistämällä sellaiset klusterit, joiden etäisyys on pienin (tai vastaavasti samankaltaisuus on suurin). Läheisyysmitta lasketaan solmuparille riippumatta siitä, ovatko nämä yhdistetyt. [Fortunato, 2010; Schaeffer, 2007] Kokoavilla menetelmillä saadut klusterit eivät välttämättä siten ole enää yhtenäisiä, varsinkaan jos käytetty läheisyysmitta ei perustu tietoon graafin rakenteesta.

Hierarkkisen klusteroinnin tulosta voidaan havainnollistaa *puukuviolla* (dendrogram) kuten kuvassa 3, jos tasojen oletetaan olevan osituksia [Fortunato, 2010]. Jos kuvaajaa luetaan alhaalta ylös, voidaan nähdä kokoavan klusteroinnin ottamat vaiheet järjestyksessä. Alin taso koostuu jokaisesta alkioista omassa klusterissaan. Seuraavalla tasolla on yhdistetty alkio 4 ja 5, joiden välinen etäisyys on kaikista alkiojokien välillä lasketuista etäisyyksistä pienin. Tällä tasolla muut alkio 1, 2 ja 3 ovat edelleen omissa klustereissaan. Ylemmällä tasolla on klusteroitu vastaavasti alkio 2 ja 3. Samoin on tehty klustereille  $\{2, 3\}$  ja  $\{3, 4\}$ , mutta tällöin klusterien välisen etäisyyden määrittämiseksi on voitu käyttää esimerkiksi yhden yhteyden menetelmää. Viimeisenä ylin vaakasuora viiva edustaa klusteria, johon kuuluvat kaikki alkio 1. Kuvassa 3 klusteri yhdistetään aina vain yhden muun klusterin kanssa tai jaetaan kahteen osaan, mutta yhden klusterin aliklustereita voisi samalla tasolla olla useampiakin kuin kaksi [Schaeffer, 2007].



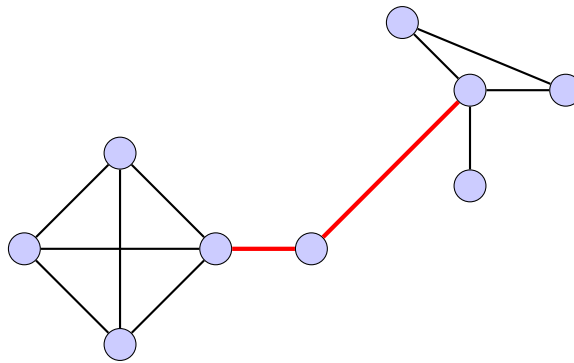
Kuva 3: Puukuvio viiden alkion klusteroinnista.

Jakavan algoritmin iteraatio- tai rekursioaskel voi perustua särmien poistamiseen edellisen askeleen tuottamista klustereista. Näin aligraafeista jää jäljelle pienempiä komponentteja, jotka edustavat uuden tason klustereita. Eräs tapa jakaa yhtenäinen graafi näin on poistaa siitä ne särmät, jotka yhdistävät sen jonkin pienen irrotuksen osajoukot toisiinsa niin, että komponentteja ovat nämä osajoukot. [Newman, 2006] Ongelmana on kuitenkin määrittää, millainen on pieni irrotus. Esimerkiksi jos valitaan *pienin* irrotus, saadaan tämä triviaalisti valitsemalla irrotuksen toinen osajoukko tyhjäksi. Tällainen tapa ei kuitenkaan anna saaduil-



le klustereille mitään mielekäästä tulkintaa. [Fortunato, 2010]. Sama ongelma on, kuinka määritellä irrotuksen osajoukkojen alkioiden lukumäärille sallittu suhde. [Schaeffer, 2007].

Tunnetuin menetelmä yhteisöjen löytämiseen on ollut Girvanin ja Newmanin [2002] jakava hierarkkinen algoritmi [Fortunato and Castellano, 2009]. Menetelmä pyrkii välttämään graafin irrotuksen käyttöön liittyviä ongelmia jakamalla graafin sen *särmien välillisyyden* (edge betweenness) perusteella. Välillisuus voidaan laskea *lyhimmän polun välillisyytenä* (shortest-path betweenness, geodesic edge betweenness). Tällöin särmälle lasketaan, kuinka monta lyhintä polkua eli geodeesista polkua graafin solmuparien välillä sisältää tämän särmän. Mitä suurempi luku särmälle saadaan, sitä paremmin särmän voidaan ajatella erottavan klustereita toisistaan; tätä on havainnollistettu kuvassa 4. Girvan ja Newman [2004] esittävät myös kaksi muuta tapaa välillisyyden laskemiseen, mutta lyhimmän polun välillisuus on näistä tavoista yksinkertaisin määritellä. Muut heidän esittämistä välillisyyksimitoista eivät myöskään yleensä johda sitä parempiin tuloksiin. [Girvan and Newman, 2004]



Kuva 4: Kaksi särmää, joilla on suurin välillisuus.

Girvanin ja Newmanin [2002] menetelmän alkuperäisessä määritelmässä ei otettu kantaa siihen, mitkä sen muodostamista hierarkian tasoista edustavat parhaimpia klusterointeja. He kuitenkin myöhemmin julkaisivat menetelmästä uuden version, jossa parhain taso valitaan sen modulaarisuuden perusteella [Girvan and Newman, 2004]. Tätä laatumittaa käsitellään seuraavaksi.

## 5.2. Modulaarisuus klusteroinnin laatumittana

Klustereiden laatua voidaan yrittää arvioida silmämääräisesti visualisoimalla aineistoa eri tavoin. Tällainen tapa voi kuitenkin käydä nopeasti hankalaksi aineiston koon kasvaessa. Toinen tapa arvioida klusterien laatua on käyttää *laatumittaa* (quality measure, quality function) eli mittaa, joka kuvaa klusteroinnin hyvyytä. Laatumittaa voidaan käyttää klusterointimenetelmien pysäytysehtona. [Fortunato, 2010]

Eri laatumitat antavat eri vastauksia sille, mitkä ominaisuudet tekevät klusterista hyvän. Tunnetuin laatumitta on ollut modulaarisuus. Alkuperäinen modulaarisuuden määritelmä perustuu oletukseen, että graafi, jossa särmät on valittu satunnaisesti solmuparien välille, ei sisällä yhteisörakennetta. Tällaista graafia kutsutaan tässä luvussa myös satunnaisgraafiksi, mutta satunnaisgraafilla voidaan tarkoittaa myös tällaisen graafin erikoistapausta. [Fortunato, 2010]

Alkuperäinen modulaarisuuden määritelmä käyttää hyväksi tietoa satunnaisgraafista, jonka solmut ja astejono ovat samat kuin siinä alkuperäisessä graafissa, jonka klusteroinnille modulaarisuus lasketaan. Satunnaisgraafin jokaisen solmun aste on siis sama kuin sitä vastaavan solmun aste alkuperäisessä graafissa. Modulaarisuus kuvaa, miten klusterien solmuihin liittyvät särmät alkuperäisessä graafissa eroavat niistä särmistä, joita samoihin solmuihin liittyy satunnaisgraafissa.

Satunnaisgraafin, jonka astejono ja solmut vastaavat alkuperäistä graafia, voidaan ajatella lähtevän muodostumaan, kun alkuperäisen graafin jokainen särmä jaetaan kahtia. Näin syntyy kaksi ”puolisärmää”, jotka molemmat sisältävät alkuperäisen särmän päätesolmuista toisen. Jos tämän graafin särmien lukumäärä on  $m$ , niin näitä puolisärmia syntyy  $2m$ . Satunnaisgraafi saadaan, kun jokainen puolisärmä yhdistetään satunnaisesti jonkin toisen puolisärmän kanssa. Jos solmujoukon  $V = \{v_1, \dots, v_n\}$  astejono on  $k_1, \dots, k_n$ , niin todennäköisyys  $p_i$  ( $1 \leq i \leq n$ ) yhdistää puolisärmä sellaisen puolisärmän kanssa, jonka päätesolmu on  $v_i$ , on  $k_i/2m$ . Täten todennäköisyys sille, että solmuja  $v_i$  ja  $v_j$  ( $1 \leq j \leq n$ ) yhdistää särmä, on todennäköisyys  $p_i p_j = k_i k_j / 4m^2$ , sillä nämä solmut valitaan riippumatta toisistaan. Odotusarvo solmuparia  $v_i, v_j$  yhdistävien särmien lukumäärälle on  $2mp_i p_j = k_i k_j / 2m$ . [Fortunato, 2010]

Modulaarisuutta määritettäessä lasketaan yhteen klusterien sisällä olevien särmien lukumäärät, joista on vähennetty särmien odotusarvo alkuperäistä graafia vastaavassa satunnaisgraafissa. Olkoot graafi  $G = (V, E)$ , solmujoukko  $V = \{v_1, \dots, v_n\}$  ja särmien lukumäärä  $|E| = m$ . Lisäksi käytetään solmun  $v_i$  ( $1 \leq i \leq n$ ) asteesta merkintää  $k_i$ . Tällöin graafin  $G$  osituksen  $\{C_1, \dots, C_l\}$  modulaarisuus  $Q(\{C_1, \dots, C_l\})$  voidaan laskea kaavalla

$$Q(\{C_1, \dots, C_l\}) = \frac{1}{2m} \sum_{v_i, v_j \in V} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j),$$

missä  $A_{ij}$  on solmujen  $v_i$  ja  $v_j$  kautta kulkevien särmien lukumäärä (lukumäärien  $A_{ij}$  sanotaan myös olevan *vierusmatriisin* (adjacency matrix)  $A$  alkioita). Merkin  $C_i$  tarkoittaa solmun  $v_i$  klusteria, eli  $v_i \in C_i$  ja  $C_i \in \{C_1, \dots, C_l\}$ . Funktio  $\delta(C_i, C_j)$  huomio sen, että lasketaan yhteen vain klusterien sisäiset särmät, ja se saa arvon 1, jos  $C_i = C_j$  ja on muuten 0. Graafi  $G$  voi sisältää sekä rinnakkaissärmia että luuppeja, mutta sen oletetaan olevan sekä suuntaamaton että painoton. [Fortunato, 2010] Kaavan ensimmäisen termin  $(1/2m)$  merkitys on pääasiassa konventionaalinen, ja

sen tarkoitus on pitää määritelmä yhteensopivana modulaarisuuden aikaisempien kaavojen kanssa [Newman, 2006]. Myös alkuperäinen modulaarisuus voidaan määrittää useilla eri tavoilla, joista tässä esitettiin yksi [Schaeffer, 2007].

## 6. Yhteenveto

Verkostoaineisto on sosiaalisen verkostanalyysin käsite, joka kuvaa havaintoyksiköitä ja näiden välisiä suhteita ominaisuusmuuttujien ja suhdemuuttujien avulla. Verkostoaineistossa on vähintään yksi suhdemuuttuja. Ne havaintoyksiköt ja suhteet, joista verkostoaineisto on mitattu, muodostavat sosiaalisen verkoston. Tällaisen verkoston käsite voidaan myös samaistaa graafiin, jonka solmuihin liittyvät verkostoaineiston ominaisuusmuuttujat ja jonka särmät kuvaavat suhdemuuttujia. Sosiaalisilla verkostoilla on erilaisia ominaispiirteitä, jotka vaikuttavat siihen, minkälaisessa graafissa verkostoaineistoa voidaan merkitä. Toisaalta verkostoaineistoja voidaan yksinkertaistaa, jotta niitä voidaan esittää yksinkertaisemmissa graafeissa.

Verkostoaineistoja voidaan luoda perinteisin tavoin kuten haastattelujen pohjalta, mutta suuria verkostoaineistoja saadaan keräämällä tietoa ihmisten välisestä viestinnästä, erityisesti sosiaalisesta mediasta. Aineisto ei sinänsä palvele välttämättä mitään tarkoitusta, ja tarve onkin saada näistä tietämystä kuten tulkintoja. Tietämystä voidaan saada tiedonlouhinnan menetelmillä osana KDD-prosessia. Sosiaalisille verkostoille tyypillinen tiedonlouhintatapa on viime aikoina ollut yhteisöjen löytäminen eli klusterointi.

Klusteroinnin tuloksena saadaan yhteisöjä eli klustereita, jotka yleisesti ottaen ovat graafien aligraafeja. Klustereille on useita eri määritelmiä. Tässä luotiin katsaus sellaisiin klusterien määritelmiin ja klustereita löytäviin menetelmiin, joita sosiaalisten verkostojen tutkimuksissa yleisimmin tavataan. Tällaisia menetelmiä ovat olleet hierarkkiset menetelmät ja erityisesti Girvanin ja Newmanin särmien välillisyyteen perustuva jakava menetelmä. Lopuksi esiteltiin, kuinka menetelmien tuottaman klusteroinnin laatua voidaan arvioida modulaarisuuden avulla.

Tämän tutkielman otsikon alla voisi käsitellä myös monia muita asioita. Sosiaalisten yhteisöjen olennaisena piirteenä nähdään esimerkiksi niiden limittäisyys ja tällaisia menetelmiä on tutkittu paljon. Tässä tutkielmassa käsitellyt klusterointimenetelmät tuottavat kuitenkin vain solmujoukon osituksia. Yhteisöjen löytämisen käsittäminen osana KDD-prosessia antaa myös laajan aiheen käsiteltäväksi. Tässä prosessissa klusterointivaihe vaikuttaa suhteellisen vähän siihen, saadanko aineistosta käyttökelpoista tietoa. Olennaisia osia sen sijaan ovat tiedon esi- ja jälkikäsitely, jotka tässä tutkielmassa jäivät vain pienelle huomiolla. Lisäksi tärkeä osa yhteisöjen löytämisestä on siinä saatujen yhteisöjen evaluointi. Tähän liittyvät kysymykset, kuten miten arvioida laatumitan laatua ja mitä näiden menetelmien tulisi lopulta tehdä.

## Viitteet

- [Fayyad *et al.*, 1996] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth, The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM* **39**, 11 (Nov. 1996), 27–34.
- [Fortunato, 2010] Santo Fortunato, Community detection in graphs. *Physics Reports* **486**, 3 (2010), 75–174.
- [Fortunato and Castellano, 2009] Santo Fortunato and Claudio Castellano, Community structure in graphs. In: Robert A. Meyers (ed.), *Encyclopedia of Complexity and Systems Science*. Springer New York, 2009, 1141–1163.
- [Girvan and Newman, 2002] M. Girvan and M. E. J. Newman, Community structure in social and biological networks. *PNAS* **99**, 12 (June 11, 2002), 7821–7826.
- [Girvan and Newman, 2004] M. Girvan and M. E. J. Newman, Finding and evaluating community structure in networks. *Phys. Rev. E* **69**, 2 (Feb. 2004), 026113.
- [Hansen *et al.*, 2011] Derek L. Hansen, Ben Shneiderman, and Marc A. Smith, *Analyzing Social Media Networks with NodeXL*. Morgan Kaufmann, Boston, 2011.
- [Hillery Jr., 1955] George A. Hillery Jr., Definitions of community: areas of agreement. *Rural Sociology* **20**, 2 (1955), 111–123.
- [Jain *et al.*, 1999] A. K. Jain, M. N. Murty, and P. J. Flynn, Data clustering: a review. *ACM Comput. Surv.* **31**, 3 (Sept. 1999), 264–323.
- [Johanson *et al.*, 1995] Jan-Erik Johanson, Mikko Mattila ja Petri Uusikylä, *Johdatus Verkostoanalyysiin, Menetelmäraportteja ja käsikirjoja*. Kuluttajatutkimuskeskus, Helsinki, 1995.
- [Knoke and Kuklinski, 1982] David Knoke and James H. Kuklinski, *Network Analysis, Sage University Papers Series*. Sage Publications, 1982.
- [Koivisto ja Niemistö, 2001] Pertti Koivisto ja Riitta Niemistö, Graafiteoriaa. Tampereen yliopisto, Matematiikan, tilastotieteen ja filosofian laitos, Moniste B 54, 2001.
- [Newman, 2003] M. E. J. Newman, The structure and function of complex networks. *SIAM Review* **45**, 2 (2003), 167–256.
- [Newman, 2006] M. E. J. Newman, Modularity and community structure in networks. *PNAS* **103**, 23 (2006), 8577–8582.
- [Papadopoulos *et al.*, 2012] Symeon Papadopoulos, Yiannis Kompatsiaris, Athena Vakali, and Ploutarchos Spyridonos, Community detection in social media. *Data Min. Knowl. Disc.* **24**, 3 (2012), 515–554.
- [Porter *et al.*, 2009] Mason A. Porter, Jukka-Pekka Onnela, and Peter J. Mucha, Communities in networks. *Notices of the AMS* **56**, 9 (2009), 1082–1097.

- [Schaeffer, 2007] Satu Elisa Schaeffer, Graph clustering. *Computer Science Review* **1**, 1 (2007), 27–64.
- [Scott, 2000] John Scott, *Social Network Analysis: A Handbook*, 2nd ed. SAGE Publications, 2000.
- [Vasudev, 2007] C. Vasudev, *Combinatorics and Graph Theory*. New Age International, Daryaganj, Delhi, 2007.

# Suomen potilastietojärjestelmien käytettävyys

**Peter Skogberg**

## Tiivistelmä.

Tässä tutkielmassa käsitellään Suomen terveydenhuollossa käytettävien potilastietojärjestelmien käytettävyyttä. Tutkielmassa kerrotaan ensin yleisellä tasolla potilastietojärjestelmistä sekä niiden käytön historiasta Suomessa. Sitten siirrytään käsittelemään Suomen potilastietojärjestelmien käytettävyyttä. Tutkielman on tarkoitus toimia koosteenä 2000-luvun puolivälin jälkeen tehdyistä potilastietojärjestelmien käytettävyystutkimuksista sekä asiaa sivuavista julkaisuista.

**Avainsanat ja -sanonnat:** Potilastietojärjestelmä, käytettävyys

## 1. Johdanto

Suomen valtauutisoinnissa kirjoitetaan ajoittain julkisista tietojärjestelmäprojekteista sekä julkisiin tietojärjestelmiin ja niiden hankintaan liittyvistä onnistumisista ja epäonnistumisista. Terveydenhuollon järjestelmät ovat kriittisiä järjestelmiä, joiden toimivuus vaikuttaa suoraan potilasturvallisuuteen sekä hoidon mahdollisimman hyvään onnistumiseen. Terveydenhuollon kustannukset muodostavat suuren osan valtion budjetista, joten hyvin toimivilla terveydenhuollon tietojärjestelmillä voitaisiin saavuttaa merkittäviä kustannussäästöjä. Mitä parempi potilastietojärjestelmän käytettävyys on, sitä paremmin se auttaa hoitotyössä.

Potilastietojärjestelmien käytettävyydestä on saatavilla kohtalaisesti suomenkielistä tutkimustietoa 2000-luvulta. Tähän työhön on pyritty kokoamaan näiden tutkimusten tuloksia sekä keräämään muistakin julkisen terveydenhuollon järjestelmistä kertovista julkaisuista potilastietojärjestelmien käytettävyteen liittyvää tietoa.

Aluksi toisessa luvussa kerrotaan hieman tarkemmin potilastietojärjestelmistä. Kolmannessa luvussa käydään läpi potilastietojärjestelmiä sekä niiden käytettävyiden tutkimusta yleisellä tasolla ja neljännessä luvussa esitellään näiden tutkimusten tuloksia, jonka jälkeen työ päättyy yhteenvetoon.

## 2. Suomen potilastietojärjestelmät

### 2.1. Potilastietojärjestelmän käyttötarkoitus

Potilastietojärjestelmän tavoitteena on helpottaa työtä ja lisätä hoitohenkilöstön aikaa potilastyöhön. Potilastietojärjestelmistä löytyy kaikki potilaasta kertynyt hoitoon liittyvä tieto siten, että se vastaa rakenteeltaan käyttötarkoitusta ja hoi-

toa. Potilastietojärjestelmiin tallennetaan potilaan hoidon kannalta oleellisia tietoja ja potilastietojärjestelmiä käyttävät terveydenhuollon ammattilaiset, kuten lääkärit, hoitajat ja vastaanottovirkailijat. Esimerkiksi röntgenkuvat ja laboratoriotutkimustulokset sekä muu potilaasta kirjattava tieto ovat välittömästi käytettävissä järjestelmästä. [2]

Potilastietojärjestelmien käytön myötä moniammatillinen yhteistyö lisääntyy, kun kirjauksia voidaan seurata reaaliaikaisesti [8]. 2000-luvun puolivälin jälkeen potilastietojärjestelmien kehityksen painopiste on siirtynyt organisaatiokeskeisestä organisaatioiden välisen toiminnan kehittämiseen [1]. Organisaatioiden välistä toimintaa ovat esimerkiksi potilastietojen sähköinen välitys, yhteiset potilastietojärjestelmät ja sähköiset arkistot.

Potilastiedot pitää saada myös tietoturvallisesti ja ymmärrettävästi terveydenhuollon ammattilaisten käyttöön eri toimintayksiköissä. Potilastietojen käytön tulee olla lainmukaista, tietoturvallista ja potilaan oikeuksia kunnioittavaa. Potilastietojärjestelmään tallennettava data ei ole homogeenista, sillä se voi koostua esimerkiksi tekstistä, kuvista sekä digitaalisista mittaustuloksista ja tietoa säilytetään pitkään; lain mukaan osin jopa pysyvästi [2].

Jauhiainen [4] mainitsee hoitotyön tärkeimmiksi tekijöiksi tietoturvan mukaisen toiminnan, kiinnostuksen ihmisestä kokonaisuudessaan, yhteistyö- ja ryhmätyövaatimukset moniammatillisessa toiminnassa, eettisen suhtautumisen työhön sekä myönteisen suhtautumisen tieto- ja viestintätekniikan käyttöön ja toteuttamiseen. Hoitotyön kehittäminen edellyttää hoitotyön tiedonhallinnan kartoittamista sekä perehdyttämistä ja kouluttamista uusiin tietojärjestelmiin [2].

Manuaalinen kirjaaminen on sisällön monipuolisuuden, päällekkäisten kirjausten ja toiston, puutteellisen täytön sekä epäselvän käsialan vuoksi heikkotasoista [13]. Sähköisen potilastietojärjestelmän on tarkoitus ratkaista esimerkiksi näitä ongelmia. Sähköinen kirjaaminen ei ole aiheuttanut kirjaamisen laadun heikkenemistä manuaaliseen kirjaamiseen verrattuna [8].

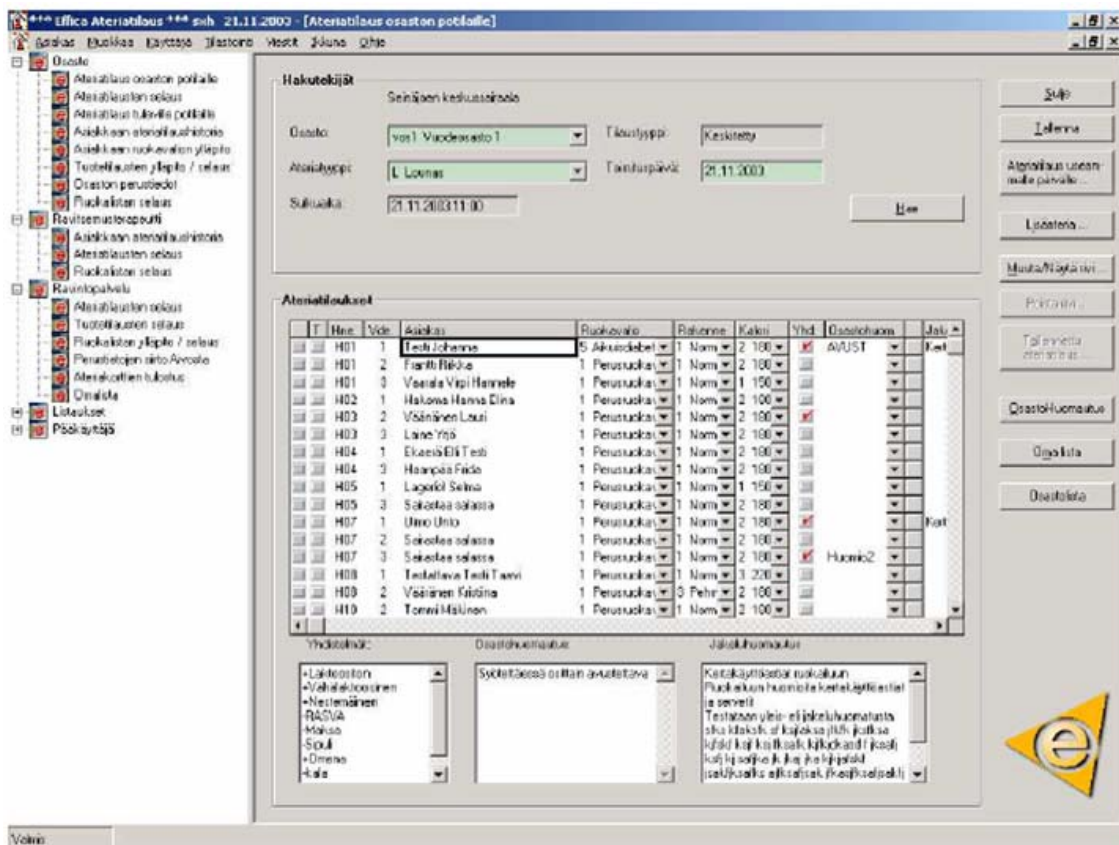
## **2.2. Suomessa käytössä olevista potilastietojärjestelmistä**

Suomen yleisimmät potilastietojärjestelmät Potilastietojärjestelmät lääkärin työvälineenä 2014 -tutkimuksen mukaan on koottu taulukkoon 1. Näiden perusjärjestelmien lisäksi on käytössä satoja erillisjärjestelmiä [14]. Lisäksi 2000-luvulla on kehitetty KanTa-nimistä resepti- ja potilastietopalvelua, joka esitellään tarkemmin kohdassa 2.4.

	Sairaala, %	Terveyskeskus, %	Yksityinen/muu, %	Kaikki, %
Acute			10,0	3,3
DynamicHealth (Doctorex)			39,6	11,7
Effica	24,5	46,1	5,0	24,1
Esko	10,6			5,4
Graafinen Finstar		4,0		1,3
Mediatri	2,7	7,7	2,7	3,9
Medicus			2,3	0,7
Pegasos	6,0	37,0	5,2	13,4
Radiologien PACS/RIS	5,2		1,1	3,0
Softmedic			10,3	3,1
Uranus	46,0		9,4	24,1
Jokin muu	5,0	5,2	14,4	6,0
Kaikki, %	100,0	100,0	100,0	100,0
n	1 753,0	922,0	1 106,0	3 781,0

Taulukko 1. Suomen pääasiallisesti käytössä olevat potilastietojärjestelmät, jossa n on kyselyyn osallistuneiden lääkäreiden lukumäärä.[14]

Tosin jotkut taulukossa 1 mainitut potilastietojärjestelmätkin ovat jäämässä hiljalleen historiaan, kun niiden toimittajat tarjoavat uusia ratkaisujaan. Tieto tarjoaa Lifecare-nimistä kokonaisratkaisua sosiaali- ja terveyshuollon yhteisiin tiedonhallintatarpeisiin. Sen tarkoitus on vähentää turhia työvaiheita ja tehostaa hoitoketjuja. Siihen tallennettua tietoa voidaan hyödyntää toiminnan resursoinnissa, kehittämisessä ja johtamisessa. Suomalaislääkäreiden kanssa yhdessä kehitetty uusi yhteenvedonäkymä on erilainen Efficaan verrattuna. [7]



Kuva 1. EfficA--potilastietojärjestelmän tukitoiminto: EfficA Ateriatilaukset [1].



CGI tarjoaa Pegasoksen ja Uranuksen sijaan uutta Desktop-palvelua, johon yhdistellään Pegasoksen ja Uranuksen toiminnallisuuksia. Sen käyttöliittymän suunnittelussa on huomioitu lääkärin työskentelyn tukeminen ja hiirenpainallusten vähentäminen siirtymien välillä. Uusi järjestelmä on tarkoitettu ottaa käyttöön ensimmäisenä yliopistosairaaloissa potilastiedon arkistoon liittymisen yhteydessä. [7]

### **2.3. Potilastietojärjestelmien käyttöönotot Suomessa**

Potilastietojärjestelmät ovat kehittyneet Suomessa voimakkaasti 2000-luvulla. Vielä vuosituhannen alussa erikoissairaanhoidon potilaskertomukset säilytettiin suurimmaksi osaksi papereilla ja vain 63% terveyskeskuksista käytti sähköistä potilaskertomusjärjestelmää [1]. Sähköisellä potilaskertomuksella tarkoitetaan potilaskertomuksen tietojen tallentamista ja hallintaa tietotekniikan avulla. Se toimii tilastoinnin tukena sekä organisaatioiden välisessä tiedonvälityksessä [2]. Vuonna 2003 käytön osuus oli kasvanut terveyskeskuksissa jo 93,6%:iin ja erikoissairaanhoidossa 62%:iin. Vuonna 2006 käytön osuus oli molemmissa jo yli 90% [1].

Vuonna 2007 säädettiin laki sähköisestä lääkemääräyksestä (L 2.2.2007/61, Laki sähköisestä lääkemääräyksestä) ja nykyään painopiste onkin siirtynyt organisaatiokeskeisestä järjestelmäkehityksestä organisaatioiden ja järjestelmien välisen viestinnän kehitykseen [1]. Organisaatioiden väliset palvelut mahdollistavat eri organisaatioiden potilastietojen kokoamisen hoitoketjun kokonaisuuden hallitsemiseksi [2].

Vuonna 2007 90%:lla sairaanhoitopiireistä oli käytössä sähköisen potilastiedon välittämiseen käytettäviä järjestelmiä [1]. Tosin silti vuonna 2014 terveyskeskuksissa 21% ja sairaaloissa 46% lääkäreistä käyttivät vielä paperia ja faksia [15]. Kehitys on kuitenkin ollut nopeaa, sillä vuonna 2010 vastaavat osuudet olivat terveyskeskuslääkäreistä 40% ja sairaalalääkäreistä 67% [14].

Käyttöönotto asettaa sekä inhimillisiä että organisatorisia haasteita [2]. Muutoksessa tulee ottaa huomioon myös potilaan hoidon jatkuvuus. Potilastietojärjestelmän toteuttaminen vaatii terveydenhuollossa koulutusta ja työtä sekä yhteistyötä käyttäjien ja tietojärjestelmän toteuttajien välillä. Tietoturva, järjestelmien yhteensovittaminen sekä yhteisen kielen löytäminen aiheuttavat haasteita [2].

Uuden potilastietojärjestelmän käyttöönoton ja kehityksen onnistumiseen auttaa hyvin organisoitu ja strateginen lähestymistapa, yksilö- ja ryhmäohjaus sekä tuki myönteiselle suhtautumiselle järjestelmää kohtaan. Yksilöohjaus on ryhmäohjausta tehokkaampaa, mutta kallista ja hankalampaa järjestää. [2]

Sosiaali- ja terveysministeriö valvoo potilastietojärjestelmien toteutusta. Myös KELA, Suomen Kuntaliitto ja TEO ovat mukana järjestelmien toteutuksessa [2].

#### **2.4. Kansallinen terveysarkisto**

Koska Suomen terveydenhuollossa on niin monia erillisjärjestelmiä, STM on suunnitellut ja toteuttanut 2000-luvun jälkipuoliskolla kansallisen tietojärjestelmäarkkitehtuurin. Tuloksena on syntynyt valtakunnallinen tietojärjestelmäpalvelu nimeltään KanTa (Kansallinen terveysarkisto). Kansallisen terveysarkiston avulla toteutetaan tiedon arkistointiin ja hallinnointiin liittyvät ominaisuudet. Siihen kuuluvat valtakunnalliset tietojärjestelmäpalvelut sähköinen resepti (eResepti), kansallinen lääketietokanta, sähköinen potilastiedon arkisto (eArkisto) sekä kansalaisten omien resepti- ja potilastietojen katselu (eKatselu).

KanTa-palvelut toteutetaan Kansaneläkelaitoksen KanTo-hankkeessa [10]. KanTa-järjestelmä toteuttaa tiedon arkistoinnin ja hallinnoinnin [1].

Palveluntuottajat sovittavat omat potilaskertomusjärjestelmänsä KanTa-palveluihin [1]. Esimerkiksi Pegasoksen kohdalla järjestelmän toimittaja CGI tarjoaa Pegasos: eArkisto-liittymää, jonka avulla Pegasos voidaan liittää lainmukaisesti KanTa-arkistoon.

KanTa-palveluiden avulla hoitohenkilökunta saa käyttöönsä tarvitsemansa potilastiedot ja apteekit saavat tarvitsemansa reseptitiedot. Lisäksi kansalaiset saavat internetin välityksellä pääsyn reseptikeskuksessa ja eArkistossa oleviin tietoihinsa [10]. Jatkossa eArkistosta saa tehtyä organisaatorajat ylittäviä hakuja [14].

Potilastiedon arkistoon tallennetaan hoidon kannalta oleellisia tietoja [5]. Tiedot tallennetaan esimerkiksi palvelutapahtuman, kuten vaikkapa osastojakson tai lääkärikäynnin aikana [5]. Potilastiedon arkistoa käytetään potilastietojärjestelmien avulla ja sen tarkoitus on mahdollistaa potilaan tietojen saatavuus yli organisaatorajojen, mikä mahdollistaa hoidon paremman jatkuvuuden ja laadun [10]. Lain mukaan tietoja saa luovuttaa eArkistosta pelkästään muille terveydenhuollon yksiköille [10]. Keväällä 2014 vain Itä-Savon sairaanhoitopiirissä oli käytössä potilastiedon arkisto [14]. Pegasos, Mediatrati ja Effica läpäisivät Kelan yhteistestauksessa 2013 auditoinnin potilastiedon arkiston suhteen [5].

Sähköinen resepti on lääkärin sähköisesti kirjaama lääkemääräys. Kirjaus tehdään Reseptikeskus-nimiseen valtakunnalliseen tietokantaan. Potilaan suostumuksella lääkäri voi tarkistaa sähköisen reseptin avulla tämän kokonaislääkityksen ja kirjata lääkemääräyksiä [10]. Sähköisen reseptin käyttö rahoitetaan käyttömaksuilla [10]. Sähköisen reseptin osalta Kelan yhteistestauksen olivat vuonna 2013 läpäisseet yleisimmistä potilastietojärjestelmistä ainakin Acute,

Effica, Esko, Mediatri, Pegasos ja Uranus [5]. Sähköistä reseptiä on pääosin keuhuttu käyttöönnoton jälkeen [14].

Täysi-ikäiset kansalaiset voivat hakea reseptikeskuksessa ja potilastiedon arkistossa olevia henkilökohtaisia hoito-, resepti- ja lokitietojaan eKatselupalvelussa. Käyttö tapahtuu internet-selaimella ja kirjautumiseen käytetään vahvaa tunnistautumista (SSL). [10]

Tammikuussa 2010 kuusi potilastietojärjestelmäklusteria katselmoitiin eReseptin ja eArkiston käyttöönottoa varten [1]. Käyttöönottojen osalta eResepti on otettu käyttöön 2010-luvulla julkisessa terveydenhuollossa ja vuonna 2013 sen käyttöasteen keskiarvo oli koko Suomessa 70% [14]. Vuonna 2014 vain Itä-Savon sairaanhoitopiiri käytti eArkistoa [14].

### 3. Potilastietojärjestelmien käytettävyydestä yleisesti

#### 3.1. Käytettävyyys käsitteenä

Käytettävyydellä mitataan sitä, miten hyvin jokin tuote auttaa käyttäjää saavuttamaan tietyt tavoitteensa tietyssä käyttökontekstissa (ISO-standardi 9241). Käyttökonteksti liittyy ympäristöön, jossa laitetta käytetään, käyttäjän ominaisuuksiin sekä tehtävänkuvaan. Hyvä käytettävyys lisää käyttäjien tyytyväisyyttä ja tehokkuutta sekä saa käyttäjän tekemään tehtävänsä nopeammin. Myös koulutusta järjestelmän käyttöön vaaditaan vähemmän, jos käytettävyys on hyvä.

Hyvää käytettävyyttä tukevat periaatteet ovat opittavuus, joustavuus ja vakaus. Ennustettava järjestelmä on johdonmukainen ja järjestelmän joustavuus mahdollistaa eri käyttötapoja, esimerkiksi monenlaisten eri syötelaitteiden käytön. Vakaa järjestelmä antaa palautetta sekä käyttäjälle mahdollisuuden korjata tehtyjä virheitä. Jos käytettävyys on hyvä, on käyttö turvallista, helposti opittavaa ja muistettavaa ja tehokasta. Siten käyttökin on tyydyttävää, viihdyttävää sekä palkitsevaa [12].

Käytettävyyteen liittyy myös läheisesti käsite käyttäjäkokemus (UX, user experience), jolla tarkoitetaan käyttäjän kokemaa yleisvaikutelmaa tuotteesta, kun käyttäjä pitelee tuotetta ensimmäistä kertaa. Käyttäjäkokemus voi olla esimerkiksi hauska, palkitseva, viihdyttävä ja motivoiva [12]. Aikaisempi kokemus tuotemerkestä, tuotteesta tai järjestelmästä sekä käyttäjän persoonallisuus ja psyykkinen tila vaikuttavat käyttäjäkokemuksen syntymiseen.

Käyttäjäkokemus muodostuu tuotteen käyttöominaisuuksista sekä käyttäjän aiemmasta kokemuksesta. Myös käyttäjän asenteet, psyykkinen tila, taidot ja käyttötilanne vaikuttavat käyttäjäkokemukseen.

Hyvän käytettävyyden saavuttamiseksi käyttäjä pitää tuntea ja ottaa mukaan tuotteen suunnitteluun. Käyttäjistä voidaan eritellä erilaisia käyttäjäprofii-leja sekä selvittää tuotteen käyttöön liittyviä rutiineja ja ongelmia esimerkiksi haastatteleamalla käyttäjiä tai tekemällä havainnointeja käyttötilanteista. Käyttä-jäkeskeiseen suunnitteluun (User-Centered Development) liittyy iteratiivinen suunnittelu, jossa olennaisessa osassa on loppukäyttäjien avustuksella tapahtu-va kehitys, prototyypit sekä monien väliversioiden toteutus lopullisen ohjelman kerralla tapahtuvan suunnittelun sijaan.

Hyvän käytettävyyden saavuttamiseksi suunnittelijoiden tulee ymmärtää käyttäjien tarpeita ja käyttötilanteita. Loppukäyttäjien ominaisuudet, tarpeet ja käyttökonteksti on selvitettävä suunnitteluvaiheessa.

### **3.2. Potilastietojärjestelmien käytettävyyden edut**

Hyvä käytettävyys on hoitotyössä tärkeää, sillä järjestelmän tiedot voivat vai-kuttaa suoraan valittuun hoitoon ja siten potilasturvallisuuteen. Käyttäjät ovat tehokkaampia ja tyytyväisempiä, kun käytettävyys otetaan huomioon järjestel-män suunnittelussa. Huono käytettävyys aiheuttaa käyttäjässä negatiivisia tun-nereaktioita ja voi johtaa tehtävän huonompaan suoritukseen. Hyvä potilastie-tojärjestelmä säästää lääkäriä toistamasta asioita, jotka järjestelmä jo tietää [14]. Hyvä käytettävyys aiheuttaa myös kustannussäästöjä [10]. Papereiden hankki-miseen kuluu vähemmän aikaa [2]. Taloudellisia menetyksiä taas saattaa ai-heuttaa järjestelmän puutteellinen käyttö, jos tärkeitä tietoja jää kirjaamatta (esim. rikos- ja tapaturmatapausten kuvaukset) [16].

Implisiittisen käyttäjän eli potilaan osalta käytettävyys näyttäytyy sujuvana hoitona [16]. Potilasta voidaan palvella tehokkaammin esimerkiksi puhelimitse tapahtuvissa palvelutilanteissa, sillä potilastiedot ovat nähtävissä välittömästi järjestelmästä [2]. Toisaalta potilas kokee ongelmalliseksi, mikäli lääkäriltä ku-luu paljon aikaa tietokoneen käyttöön ja potilas saattaa tällöin esimerkiksi kes-keittää hetkellisesti sanallisen kuvauksensa oireistaan [16].

Yleisellä tasolla potilastietojen tallentamisen pääasialliset hyödyt ovat tie-don parantunut luotettavuus ja saatavuus. Maantieteellisesti eri sijainneissa työskentelevät terveydenhuollon toimijat pääsevät käsiksi samoihin tietoihin. Hyvin toimivien potilastietojärjestelmien katsotaan myös olevan edellytys sote-uudistuksen onnistumiselle [14]. Potilastietojärjestelmät helpottavat pitkän ai-kavälin potilastietojen keräämisessä ja tiedoista voidaan koostaa esimerkiksi raportteja [6]. Esimerkiksi Effica Raportointi yhdistelee tietoja terveydenhuollon toiminnasta erikoissairaanhoidosta, perusterveydenhuollosta sekä erityishuol-topiireistä (esim. työtoiminta, perhehoito, laitoshoido) [3]. Raportoinnin hyö-dyiksi lasketaan esimerkiksi parantunut tietojen yhdisteltävyys, kun läheteistä,

diagnooseista ja toimenpiteistä sekä laskutuksesta voidaan koostaa tietoa samalle raportille [3].

### **3.3. Mitkä asiat vaikuttavat potilastietojärjestelmien käytettävyyteen?**

Käyttäjien perehdyttäminen ja koulutus järjestelmän käyttöön on tärkeää [6]. Ohjelmistojen joustavuus ja toimivuus tietyssä käyttökontekstissa ovat merkittäviä ominaisuuksia ja helpottavat kirjaamisten tekemistä [2]. Walldénin [16] käyttämässä teoreettisessa viitekehyksessä käytettävyyden määritellään koostuvan Nielsenin käytettävyystekijöiden mukaisesti käyttöliittymän opittavuudesta, tehokkuudesta, muistettavuudesta, virheettömyydestä ja miellyttävyydestä. Samanlaisuus ja johdonmukaisuus järjestelmän eri osissa vaikuttavat positiivisesti opittavuuteen, samoin käyttöohjeet ja järjestelmän tarjoama palaute [2].

Tehokkuudella tarkoitetaan käytön nopeutta kokeneemmilla käyttäjillä. Muistettavuus mittaa sitä, miten helposti kokeneemmat käyttäjät muistavat toimintojen, termien ja graafisten merkkien sisällön. Järjestelmiä pitäisi pyrkiä kehittämään lääkäreiden kanssa [14]. Järjestelmien käyttötilanteet myös eroavat toisistaan. Esimerkiksi murtumalääkäreiden kohdalla järjestelmän käyttö on suppeampaa, sillä hoitoon hakeutumisen syyt ovat yleensä samankaltaisia, kun taas vaikkapa päivystyksessä on hyvinkin erilaisia hoitotilanteita [16].

Tieto- ja viestintäteknikan käyttö eroaa eri organisaatioiden välillä. Erillistietojärjestelmien runsaus koetaan ongelmana. Henkilökunta tarvitsee koulutusta sekä perehdytystä järjestelmien käyttöön. Koulutuksen järjestäminen riittävillä resursseilla sekä johtajien positiivinen asenne koetaan auttavan kirjaamisten onnistumisessa [2]. Esimerkiksi Walldénin tutkimuksessa mukana olleet lääkärit (päivystysaseman päivystäjä ja murtumalääkäri) oli opastettu potilastietojärjestelmien käyttöön kaksipäiväisellä koulutuksella. Potilaiden käyntien syyt myös vaihtelevat, joten järjestelmää käytetään monipuolisesti [16].

Potilastietojärjestelmien käytettävyyssurainta tehdään kansallisella tasolla maailmanlaajuisesti vähän. Esimerkiksi Tanskassa tehdään käyttökartoituksia, jotka eivät kuitenkaan pääosin keskity käytettävyyteen ja Yhdysvalloissa kyselyt tehdään pääosin hankinnan tueksi ja käytettävyyttä jää vähemmälle huomiolle. [15]

## **4. Suomen potilastietojärjestelmien käytettävyys**

### **4.1. Miten potilastietojärjestelmien käytettävyyttä on tutkittu?**

Honkimaa-Salmi [2] esittelee SWOT-analyysin avulla hoitohenkilökunnan kokemuksia sähköisten potilastietojärjestelmien tulosta terveydenhuoltoon. Tut-

kielman teoreettisena viitekehyksenä toimineessa SWOT-analyysissä tarkasteltiin potilastietojärjestelmän uhkia, heikkouksia, mahdollisuuksia ja vahvuuksia. Aineisto kerättiin järjestämällä potilastietojärjestelmien käyttäjille teemahaastatteluita.

Honkimaa-Salmen haastattelemat henkilöt valittiin siten, että otanta oli mahdollisimman monipuolinen ja tutkija ei voinut itse valita henkilöitä ja siten vaikuttaa tuloksiin. Osallistujat olivat sairaanhoitajia vuodeosastoilta, poliklinikoilta sekä teho-, tarkkailu- ja toimenpideyksiköistä. Haastattelut nauhoitettiin ja kirjattiin kokonaisuudessaan kirjallisiksi dokumenteiksi [2]. Tämän jälkeen aineisto luokiteltiin siten, että hyvät puolet luokiteltiin joko mahdollisuuksiksi tai vahvuuksiksi ja huonot puolet luokiteltiin joko uhkiksi tai heikkouksiksi.

Walldén [16] pohti potilastietojärjestelmien käytettävyydestä tutkimuksen käytäntöjä, käytettävyydestä tutkimusmetodeja ja järjestelmien käyttökonteksteja. Tutkimuksessa on pyritty painottamaan käyttökontekstin merkitystä ja ottamaan huomioon myös potilas implisiittisenä käyttäjänä. Tutkimuksen viitekehys koostuu käyttäjistä, käyttökontekstista, Nielsenin viidestä käytettävyydestä (opittavuus, tehokkuus, muistettavuus, virhetilanteet, tyytyväisyys) ja käytettävyyden vaikutuksesta potilaaseen.

Arviointikohteena Walldénin tutkimuksessa oli laaja potilastietojärjestelmä, jonka käytettävyyttä läpikäytiin ajanvarauksen, laskutuksen, päivystysjonon, lähete- ja palautejärjestelmän sekä potilaskertomuksen osalta. Tutkimuksessa keskityttiin murtumapotilaan hoitoketjuun. Murtumapotilaat ovat heterogeeninen ryhmä, jonka oireet ovat kuitenkin melko samankaltaisia.

Aineisto Walldénin tutkimukseen kerättiin kyselylomakkeella havainnoinnilla, heuristisella arvioinnilla, käytettävyydestä ja teemahaastattelulla. Kysymyslomakkeella ja teemahaastattelulla kerättiin tietoa siitä, mistä hyvä käytettävyys potilastietojärjestelmässä käyttäjien mielestä muodostuu. Kyselylomake täytettiin tutkijan läsnäollessa. Kysymykset sisälsivät sekä avoimia että valmiita vastausvaihtoehtoja.

Havainnoinnissa valokuvattiin käyttöympäristöä. Kyselylomakkeen ja havainnoinnin avulla kerättyä aineistoa täsmennettiin teemahaastattelussa, joissa aiheina olivat myös mm. käytettävyyden vaikutukset potilaan saamaan hoitoon sekä järjestelmän vahvuudet ja heikkoudet. Kahdelle käyttäjälle tehtiin myös käytettävyydestä, jonka avulla oli tarkoitus selvittää potilastietojärjestelmien kehittämistarpeita käytettävyyden näkökulmasta. Käytettävyydestä sisälsi havainnointia, testitehtäviä, haastatteluita sekä ääneenajattelua.

Potilastietojärjestelmien käytettävyyttä on kartoitettu valtakunnallisesti vuosina 2010 ja 2014. Vuonna 2014 toteutettiin Potilastietojärjestelmät lääkärintyövälineenä-tutkimus Lääkäriliiton, THL:n, Aalto-yliopiston ja Oulun yliopis-

ton toimesta. Tutkimusta tuki taloudellisesti Sosiaali- ja terveysministeriö. Tutkimuksen tavoitteena oli saada potilastyötä tekeviltä lääkäreiltä ajantasaista tietoa järjestelmien käyttökokemuksista sekä selvittää niiden keskeisimpiä ongelmia ja hyviä ominaisuuksia.

Molemmilla kerroilla tutkimusten menetelmänä oli sähköinen kyselytutkimus, jossa pääosin arvioitiin väittämiä 5-luokkaisilla Likert-asteikoilla. Kysymysten aihealueita olivat käytettävyys, potilasturvallisuus, tekniset ominaisuudet, hoidon laatu sekä yhteistyö ja tiedonkulku. Vuonna 2014 kysyttiin lisäksi sähköisestä reseptistä. Kyselylomake testattiin etukäteen ja tutkimukseen osallistuvat 3781 työkäistä sekä potilastyötä tekevää lääkäriä kerättiin Lääkäriliiton rekisteristä lähettämällä tutkimusviestit sähköpostitse. Tutkimuksen voidaan katsoa edustavan hyvin kohdejoukkoa ja toimivan hyvänä perustana sekä normiston että yksittäisen tietojärjestelmän kehitystyössä. [14]

Suomen käytössä olevat potilastietojärjestelmät arvioitiin 2010 ja 2014 myös merkeittäin. Vuonna 2014 tarkasteluun otettiin yli 25 vastausta saaneet potilastietojärjestelmät. Vertailussa valittiin kaksi teknisiin ominaisuuksiin liittyvää asenneväittämää ja lisäksi 12 käytettävyyteen, kuusi tiedon esitystapaan, viisi potilasturvallisuuteen ja hoidon laatuun, kolme lääkitykseen sekä 10 yhteistyöhön ja tiedon kulkuun liittyvää väittämää. [15]

#### **4.2. Potilastietojärjestelmien käyttöönottoon liittyvät mahdollisuudet, uhat, vahvuudet ja heikkoudet**

Honkimaa-Salmen [2] sairaanhoitajille 2004 kohdistamassa tutkimuksessa käyttökokemus sähköisistä järjestelmistä vaikutti siihen, koettiinko järjestelmien käyttöönotto myönteisenä. Käyttäjät arvostivat mahdollisuutena tiedon helppoa saatavuutta ja siirrettävyyttä organisaatioiden välillä. Uuden oppiminen koettiin hankalaksi, mutta järjestelmän mahdollisuuksiin uskottiin. Sähköisen järjestelmän koetaan helpottavan hoitotyötä, mutta käytön opettelu vie aikaa. Tietojen uskottiin löytyvän järjestelmästä helposti ja tiedon luettavuutta sekä nopeaa saatavuutta pidettiin vahvuuksina. Hoidon laadun uskottiin paranevan työn nopeutuessa. Potilaan tietosuojaan uskottiin paranevan, sillä dokumentteihin tutustumisesta jää merkintä järjestelmään. Toisaalta mainittiin, että potilaat eivät välttämättä halua tietojensa siirtoa muihin organisaatioihin. Dokumentoinnin katsottiin parantuvan kirjaamisen selkiytyessä ja yksinkertaistuessa, kun kirjaaminen rakenteistuu, kirjaamisen läpinäkyvyys paranee ja useasti toistettavaa työtä poistuu. Raportoinnin koettiin tehostuvan suullisen raportoinnin nopeutuessa ja pääpainon siirtyessä kirjalliselle raportoinnille.[2]

Potilastietojärjestelmien heikkouksiksi mainittiin tekniikan ja työtilojen riittämättömyys. Koulutus ja ohjaus mainittiin aloitetun liian aikaisin varsinaiseen

järjestelmään siirtymiseen nähden. Tärkeänä pidettiin, että erilaisia tukihenkilöitä nimetään yksikköihin. Hoitajat kokivat olevansa käyttäjiä, eivät asiantuntijoita ja halusivat siten teknisten ongelmien selvityksen tietojärjestelmäammattilaisten tehtäväksi. Siirtymävaiheessa työmäärä lisääntyy, kun järjestelmän käyttöä pitää opetella. Siirtymävaihe aiheutti kielteisiä tuntemuksia, kuten stressiä ja ahdistusta. Tekniikasta kokonaan syrjäytyneet jäivät järjestelmän mahdollisuuksien ulkopuolelle. Heikkoutena pidettiin myös käyttäjien vaihtuessa tapahtuvan tietokoneen uudelleenkäynnistämiseen kuluva ylimääräistä aikaa. Hoidon laadun koettiin mahdollisesti heikkenevän, sillä hoitajilta vaaditaan atk-osaamista ja tietojärjestelmän käyttö vie aikaa pois potilaalta. [2]

Etätoiminta sekä samoihin tietoihin perehtyminen eri toimipisteissä koettiin mahdollisuuksina. Hoidon laadun odotettiin parantuvan turvallisuuden lisääntyessä suurten paperimäärien vaihtuessa sähköisiin kaavakkeisiin. [2]

Uhkina pidettiin toimintaongelmia, joidenka aikana tietoja ei saataisi ja täten varasysteemit koettiin välttämättöminä. Myös virukset, tietomurrot ja tietojen päätyminen väärin käsiin koettiin uhkiksi. Hoitotyön pelättiin myös teknistyvän liikaa. [2]

#### **4.3. Potilastietojärjestelmien käyttökontekstin merkitys**

Walldén [16] käsittelee käyttökontekstin vaatimuksia potilastietojärjestelmälle. Käyttäjät profiloitiin ja segmentoitiin ammattiroolien perusteella lääkäreihin, sairaanhoitajiin ja vastaanottovirkailijoihin.

Järjestelmän käyttäjinä toimivat myös sairaanhoitajat päivystysasemalla ja murtumaklinikalla. Murtumaklinikan sairaanhoitajilla järjestelmän käyttökonteksti vaihtelee vain vähän ja päivystysaseman sairaanhoitajilla paljon. Lisäksi sairaanhoitajat käyttävät vastaanoton työlistaa sekä hoitavat osan vastaanottovirkailijan työstä. Kolmas käyttäjäryhmä, vastaanottovirkailijat, hoitavat ilmoittautumisen ja laskutuksen sekä päättävät potilaan hoitoonohjauksesta. [16]

Eri tutkimusmenetelmien havaittiin tukevan toisiaan. Esimerkiksi käytettävyysskyselyssä löydetty käytettävyyssongelma saattoi havainnoinnin yhteydessä osoittautuakin työkuultuuriin liittyväksi tekijäksi. Arvioitavia käyttökonteksteja murtumapotilaan hoidossa ovat päivystysaseman vastaanottopiste, lääkärin vastaanotto sekä murtumaklinikan sairaanhoitajan vastaanotto. [16]

Käyttökontekstitekijät jaoteltiin neljän eri jäsennysmallin mukaisesti. Vuorovaikutusmallissa havaittiin, että joissain hoitotilanteissa lääkäreiden ja sairaanhoitajien täytyy neuvotella tietyistä hoitoon liittyvistä asioista, jotka pystytäisiin kirjaamaan myös potilastietojärjestelmään. Sekvenssimallilla tarkasteltaessa haasteiksi havaittiin sairaanhoitajien ja lääkäreiden työn keskeyttävät puhelinoimit, kiiretilanteet sekä se, että hoitajat joutuvat kirjautumaan järjestel-



mään pahimmillaan useita kertoja tunnissa. Joskus samat tiedot kirjataan myös useita kertoja potilaskertomuksen eri näkymiin; esimerkiksi sairaanhoitaja ja lääkäri saattavat kirjoittaa samat asiat potilaasta omiin näkymiinsä. [16]

Kulttuurimallissa tulostavasta johtuvan päivystystoiminnan ulkoistuksen koetaan aiheuttavan ongelmia, sillä alati vaihtuvien työntekijöiden järjestelmäosaaminen vaihtelee ja erityistilanteiden hallinta on heikompaa. Päivystävien lääkäreiden käyttötaito vaihtelee. Sairaanhoitajat joutuvat perehdyttämään tilapäisiä työntekijöitä järjestelmän käytössä. Laitteita myös päivitetään ja huolletaan, mikä aiheuttaa käyttökatkoksia ja sitä kautta vaikeuttaa hoitotoimenpiteitä. Fyysisen mallin osalta mainitaan tietokoneen usein sijaitsevan siten, että sen käyttö estää katsekontaktin säilyttämisen potilaaseen. [16]

Kyselylomakkeen vastauksista selvisi myös, että kiireen vuoksi tehokkuutta pidettiin tärkeänä. Potilaille aiheutuvien vaaratilanteiden syiksi mainitaan esimerkiksi järjestelmäviiveet, sekä tiedon pirstaloituneisuuteen liittyvät ongelmat (esim. monisairaiden potilaiden kaikki tiedot eivät löydy). Käyttökonteksti vaikutti siihen, mitkä tekijät käyttäjät kokivat tärkeimmiksi käytettävyyden suhteen ja käytettävyyden taso vaihteli eri käyttäjäryhmien välillä. [16]

#### **4.4. Potilastietojärjestelmien käytön miellyttävyys**

Walldénin [16] tutkimuksen mukaan säännöllisestä käytöstä johtuen miellyttävyys koettiin myös tärkeäksi ja järjestelmässä tapahtuvan toistuvuuden takia opittavuus nostettiin myös esiin. Käyttöä ohjeistavia opasteita koettiin olevan liian vähän ja terminologian ei aina katsottu olevan tunnistettavaa. Käyttäjän piti myös tietyissä tilanteissa syöttää jo edelliselle näytölle syöttämiään tietoja uudelleen. Tämä aiheuttaa lisävaivaa ja voi lisätä virheitä. Tietyt perusrutiinit myös vaativat usean ikkunan avaamisen. Päivystävät lääkärit joutuvat myös kiireen takia käyttämään sanelun tai järjestelmän käytön sijaan muistilappuja, joista tieto myöhemmin siirretään potilaskertomukseen.

Miellyttävyydestä ja muistettavuudesta annettiin hyviä arvioita. Käyttäjät pitivät myös ei-sähköisissä sairauskansioissa jo ennestään esiintyvistä värikoodi-metaforista. [16]

Päivystysaseman vastaanottovirkailijan tehtävissä käytettävyys on hyvä, kun taas sairaanhoitajien järjestelmästä löydettiin vakavia käytettävyyso ongelmia. Lääkäreiden näkökulmasta käytettävyys oli tyydyttävää. Eri metodeilla saadut tulokset olivat osin keskenään ristiriitaisia; heuristisessa läpikäynnissä löydettiin tietyn käyttäjäryhmän kohdalla paljon käytettävyyso ongelmia, joita ei havainnoimalla ja testeissä havaittu. [16]

#### 4.5. Potilastietojärjestelmien käytettävyys tuotemerkeittäin

Vänskän ja muiden selvityksessä [15] terveystietojärjestelmistä Mediatri ja Graafinen Finstar (GFS) saivat lääkäreiltä keskimääräistä paremmat arvosanat ja eniten kiitettäviä. Sairaalejärjestelmistä korkeimmat keskiarvot saivat Esko ja Mediatri ja matalimmat Uranus ja Pegasos. Eskon ja Mediatriin kohdalla tuloksiin voi vaikuttaa käyttäjien osallistuminen järjestelmien suunnitteluun.

Vuoden 2010 tutkimuksesta GFS, Pegasos ja Mediatri nousivat arvioissa, kun taas Effica laski. Kolme neljäsosaa lääkäreistä kritisoi toisesta organisaatiosta kirjatun tiedon saamisen olevan hidasta ja kolmasosa ilmoitti järjestelmävirheen vaarantaneen potilasturvallisuuden. [15]

Rutiinitehtävien sujuvuuden katsottiin tutkimuksessa kehittyneen jopa negatiivista kehitystä, mutta tämä saattaa myös johtua hoitotilanteesta tai siitä, että työprosessi on jätetty uudistamatta otettaessa uutta sähköistä työvälinettä käyttöön. Yksityinen/muu-ryhmän lääkärit pitivät rutiinitehtävien suorittamista suoraviivaisempana, kuin sairaaloiden ja terveystietojärjestelmien lääkärit. [14]

Kuntasektorilla lääkäreiden antama keskiarvo järjestelmille oli 6,5 ja yksityisellä sektorilla 7,5. Kiitettäviä arvostajia oli harvassa ja joka viides lääkäri antoi kouluarvosanaksi korkeintaan 5. Lääkärin ikä, käyttökokemus ja erikoisala vaikuttivat arvostukseen. Pidempi käyttökokemus vaikutti positiivisesti tuloksiin ja päinvastoin. Vastaajista yli 70% arvioivat olevansa kokeneita järjestelmän käyttäjiä ja vain kymmenesosalla vastaajista oli alle vuoden kokemus arvioimastaan järjestelmästä. Joka toinen vastaajista käytti säännöllisesti järjestelmää usealta eri tuotemerkiltä, kun taas reilu viidesosa käytti vain yhtä tuotemerkkiä. Yli kuusi vuotta järjestelmiä käyttäneet antoivat parhaat arvosanat. Erikoislääkäreistä huonoimmat arvostelut antoivat kirurgit, psykiatrit ja sisätautilääkärit, kun taas radiologit antoivat keskiarvoksi yli 7. [14]

#### 4.6. Terveystietojärjestelmien käytettävyys tuotemerkeittäin

Vuoden 2014 Potilastietojärjestelmät lääkärin työvälineenä -tutkimuksessa GFS:n (Graafinen Finstar) käyttäjät olivat tyytyväisimpiä tekniseen vakauteen ja reagointinopeuteen, kun taas Effican ja Pegasoksen käyttäjät toivat useimmin esiin järjestelmän hitauden ja yllättävät käyttökatkot kehittämiskohteina. Molemmilla tutkimuskerroilla (2010 ja 2014) GFS:n käyttäjät antoivat myönteisimmät ja Pegasoksen käyttäjät kielteisimmät käytettävyysarviot. Pegasoksen ja Effican käyttäjistä vain viidennes piti rutiinitehtävien suorittamista onnistuneena, kun taas GFS:n ja Mediatriin käyttäjistä puolet oli tätä mieltä. Vuonna 2014 aikaisempaa suurempi osa vastaajista koki järjestelmien vaativan paljon opettelua. Kuvantamis- ja laboratoriotutkimusten kerrottiin olevan suuren osan vas-

taajista mielestä helposti saatavilla, mutta kaksi kolmasosaa käyttäjistä oli sitä mieltä, että yhteenvetonäkymät eivät olleet kattavia. [15]

Puolet vastaajista koki järjestelmien estävän lääkitysvirheiden syntymistä ja GFS:n ja Mediatriin lääkelistoja pidettiin selkeimpinä. Muiden organisaatioiden lääketietojen koettiin olevan hyvin saatavilla kolmasosan mielestä Mediatriin käyttäjistä, kun muilla järjestelmillä osuus oli vain kymmenesosa. Potilastietojen saaminen organisaatorajojen yli koettiin edelleen liian hitaana. Sähköinen resepti mainittiin hyvänä toiminnallisuutena muuten, mutta Effican käyttäjien mielestä sähköisen reseptin käytettävyyttä pitäisi korjata. [15]

#### **4.7. Sairaalahjärjestelmien käytettävyys tuotemerkeittäin**

Teknisesti vakaimmaksi ja nopeimmaksi valittiin Vänskän ja muiden selvityksessä [15] 010 ja 2014 Esko, kun taas Uranuksen ja Effican arviot olivat huonontuneet reagointinopeuden osalta. Eskon arvioit käytettävyydestä olivat positiivisimmat ja Pegasoksen ja Uranuksen negatiivisimmat. Mediatriin arviot parantuivat eniten. Eskoa käyttävistä lääkäreistä puolet piti rutiinitehtävien suorittamista suoraviivaisena, kun taas muiden järjestelmien käyttäjistä 17-38%. Tietojen häviämistä koki enää 23-34% kaikista vastaajista.

Eskon käyttäjistä 27% oli sitä mieltä, että järjestelmä vaatii pitkää perehdytystä, muiden järjestelmien käyttäjistä tämä osuus oli puolet. Yhteenvetonäkyymiin tyytyväisiä olivat kolmasosa käyttäjistä, paitsi Pegasoksen kohdalla tyytyväisten osuus oli reilu kymmenesosa vastaajista. Vastaajista Pegasoksen, Mediatriin ja Eskon käyttäjät pitivät yhteenvetonäkymien kehittämistä tärkeänä kehittämiskohteenä. Puolet muiden kuin Eskon (15%) käyttäjistä vastasi järjestelmän olleen lähellä aiheuttaa tai aiheuttaneen potilaalle vakavan haittatapahtuman. [15]

Mediatriin käyttäjistä yli puolet vastasi tiedon muissa organisaatioissa määrättyistä lääkkeistä olevan helposti saatavilla, kun taas muiden järjestelmien käyttäjistä vain kymmenesosa oli tätä mieltä. Mediatriin, Pegasoksen ja Uranuksen osalta potilaan lääkelistan koettiin parantuneen vuodesta 2010. Interaktio mainittiin yleisesti hyvänä ominaisuutena. Myös sairaalahjärjestelmien kohdalla sähköisen reseptin käytettävyys nousi esiin Effican kohdalla, samoin Eskon käyttäjät mainitsivat sen kehityskohteenä. Uranuksen, Mediatriin ja Pegasoksen käyttäjistä suuri osa piti sähköistä reseptiä hyvänä ominaisuutena. Vajaa puolet lääkäreistä koki järjestelmien tukevan yhteistyötä hoitajien kanssa. [15]

#### **4.8. Eri organisaatioiden välinen tiedonkulku**

Organisaatioiden välinen tiedonkulku koettiin huonoksi muuten, paitsi Mediatriin kohdalla. Noin kaksi kolmasosaa lääkäreistä koki järjestelmät hyödyllisiksi samassa organisaatiossa toimivien lääkäreiden välisessä tiedonkulussa.

Puolet koki niiden tukevan sairaanhoitajien ja lääkäreiden välistä tiedonkulkua (vrt. 43% v. 2010). Vain kymmenesosa lääkäreistä koki järjestelmät hyödyllisiksi potilaiden ja lääkäreiden välisessä yhteydenpidossa ja vain pieni osa terveydenhuollon ja potilaiden välisistä yhteydenotoista tapahtuu sähköisesti. Potilaan ja terveydenhuollon sähköisen yhteydenpidon katsotaan lisääntyvän jatkossa, kun potilaat pääsevät katsomaan terveystietojaan ja kirjaamaan omia tietojaan. [14]

Potilastiedon saatavuuden lääkärin sijainnista riippumatta katsottiin olevan hyvä, samoin digitaalisesta sanelusta annettiin hyvää palautetta. Suurimmat kehityskohteet ovat vakauden, nopeuden ja eri organisaatioiden välisen tiedonkulun parantaminen. [14]

## 5. Yhteenveto

Työn tarkoitus oli esitellä Suomen potilastietojärjestelmien käytettävyyttä. Todettiin, että hyvän käytettävyyden omaava tuote tai palvelu täyttää käyttötarkoituksensa tehokkaasti ja tyydyttävästi käyttäjälleen. Jotta tuotteesta saadaan suunniteltua käytettävä, on käyttäjä tunnettava ja käyttäjää tulee kuunnella eri suunnitteluvaiheissa.

Potilastietojärjestelmien tarkoitukseksi todettiin hoitoon liittyvien tietojen tallentamisen ja käsittelyn helpottaminen. Potilastietojärjestelmien avulla saadaan myös koostettua raportteja. Hyvin toimiva potilastietojärjestelmä nopeuttaa hoitoa ja parantaa turvallisuutta sekä vapauttaa hoitohenkilökunnan työaika itse potilastyöhön. Myös potilaat saavat tietonsa helpommin käyttöönsä.

Potilastietojärjestelmien kehitykseen ja käyttöön vaikuttaa oma erityinen lainsäädäntönsä ja niiden kehitystä valvoo Sosiaali- ja terveysministeriö. Yleisiä Suomessa käytössä olevia potilastietojärjestelmiä ovat esimerkiksi Esko, Effica, Uranus ja Mediatri. 2000-luvun puolivälin jälkeen näiden järjestelmien tiettyjä tietoja on pyritty kokoamaan Kansallinen terveystietojärjestelmä-nimiseen palveluun, joka pitää sisällään reseptien käsittelyyn, potilastiedon käsittelyyn ja arkistointiin sekä potilaan omien tietojen tarkistamiseen liittyviä toiminnallisuuksia.

Potilastietojärjestelmiä on kehitetty ja otettu käyttöön Suomessa voimakkaasti 2000-luvulla. Niiden käytön osuus on terveyskeskuksissa ja erikoissairaanhoidossa yli 90%. Nykyään painopiste kehityksessä on organisaatioiden välisessä kommunikaatiossa. Käyttöönotto aiheuttaa haasteita, kun hoitotyön lisäksi pitää opetella järjestelmän käyttöä.

Jos potilastietojärjestelmässä on hyvä käytettävyys, säästää se käyttäjien aikaa ja aiheuttaa kustannussäästöjä. Tiedon luotettavuus ja saatavuus paranee. Käyttäjien koulutus koettiin tärkeäksi hyvän käytettävyyden saavuttamiseksi ja

järjestelmiä pitäisi pyrkiä suunnittelemaan lääkäreiden, sairaanhoitajien sekä muiden käyttäjäryhmien kanssa. Käyttökonteksti vaihtelee eri terveydenhuollon yksiköiden ja käyttäjien välillä paljon, joten potilastietojärjestelmiä käytetään monipuolisesti.

Työssä esiteltiin kolme suomalaista käytettävyystudkimusta vuosilta 2004, 2009 ja 2014. Käytettävyystudkimuksia on toteutettu mm. teemahaastatteluin, havainnoimalla ja käytettävyystesteillä. Haastatteluista on tehty sekä sairaanhoitajille että lääkäreille.

Honkimaa-Salmen vuonna 2004 toteuttamassa SWOT-analyysissä sairaanhoitajat uskoivat järjestelmän mahdollisuuksiin käyttöönottovaiheessa. Tiedon saatavuuden ja kirjaamisen uskottiin selkiytyvän ja päällekkäistä työtä poistuvan. Raportoinnin parantuminen nähtiin mahdollisuutena. Heikkoutena pelättiin työtilojen ja teknisten apuvälineiden riittämättömyyttä, puutteellista koulutusta, potilaan kanssa vietetyn ajan vähenemistä ja työmäärän lisääntymistä. Toimintaongelmia pidettiin uhkana.

Työssä esiteltiin Walldénin tutkimus vuodelta 2009, jossa keskityttiin potilastietojärjestelmien käytettävyyden tutkimusmenetelmiin sekä käyttökontekstiin liittyviin tekijöihin. Monen samanaikaisesti käytössä olevien tutkimusmenetelmien havaittiin tukevan toisiaan. Käyttökontekstiin liittyviä haasteita ovat esimerkiksi kesken hoidon tapahtuvat puhelinsoitot, kiire sekä ulkoistettujen työntekijöiden vaihteleva käyttötaso. Samoja tietoja saatetaan kirjata eri työntekijöiden toimesta turhaan useita kertoja. Tietokonepäätteet saattavat sijaita hoidon kannalta huonoissa paikoissa ja niiden huolto aiheuttaa haasteita.

Walldénin tutkimuksesta esiteltiin myös käytön tehokkuuteen ja miellyttävyyteen liittyviä tekijöitä. Kiireen takia lääkärit käyttivät muistilappuja ja tietyt perusrutiinit vaativat useiden ikkunoiden avaamisen. Miellyttävyydestä ja muistettavuudesta annettiin hyviä arvioita. Eri käyttäjäryhmillä käytettävyyks vaihteli.

Sairaala- ja terveyskeskusjärjestelmät arvioitiin merkeittäin lääkäreille vuonna 2014 tehdyn kyselytutkimuksen perusteella. Terveyskeskusjärjestelmien kohdalla Graafinen Finstar sai parhaat käytettävyyssarviot ja Pegasos huonoimmat. Sairaalahjärjestelmien kohdalla Eskon arviot käytettävyydestä olivat positiivisimmat ja Pegasoksen ja Uranuksen negatiivisimmat. Järjestelmistä arvioitiin esimerkiksi tekninen vakaus, tehtävien onnistuminen ja käytön opetteluun vaativuus. Järjestelmien koettiin vaativan paljon perehtymistä ja opettelua. Organisaatioiden välinen tiedonkulku koettiin ongelmalliseksi.

## Viiteluettelo

- [1] Honkanen Petri, *Sähköisen potilastietojärjestelmän hankintaprosessin arviointi*, Tampereen teknillinen yliopisto, tieto- ja sähkötekniikan tiedekunta, diplomityö, Tampere, 2010.
- [2] Honkimaa-Salmi Anu, *Sähköisen potilastietojärjestelmän tulo terveydenhuoltoon hoitohenkilökunnan kokemana*. Pro-gradu-tutkielma, Hoitotieteen laitos, Tampereen yliopisto, 2006.
- [3] Hyvönen Johanna, Kuivalainen Aki ja Pikkarainen Hannu, *Yhtenäinen raportointiratkaisu sosiaali- ja terveydenhuollon Efficaan*, Luentokalvot, Terveydenhuollon ATK-päivät 25.5.2010, Tampere.
- [4] Jauhiainen Annikki, *Tieto- ja viestintätekniikka tulevaisuuden hoitotyössä*, Kuopion yliopisto, Hoitotieteen laitos, Väitöskirja, Kuopio, 2004.
- [5] Jormanainen Vesa, *Potilastietojärjestelmät*, THL, STHJ5-kurssin luentokalvot, Hjelt-instituutti, Helsinki 2013  
<http://dspace2.lib.helsinki.fi:8082/dikk/bitstream/handle/2455/138874/STHJ2013-L1-Jormanainen.pdf>.
- [6] Kay Misha, Santos Jonathan and Takane Marina, *Management of Patient Information: Trends and Challenges in Member States*, WHO 2012.  
[http://apps.who.int/iris/bitstream/10665/76794/1/9789241504645\\_eng.pdf](http://apps.who.int/iris/bitstream/10665/76794/1/9789241504645_eng.pdf)
- [7] Kuittinen Teppo, Effica, Pegasos ja Uranus jäävät historiaan, *Mediuutiset*, 11.3.2014.
- [8] Leino Kaija, *Hoitotyön kirjaamisen toteutuminen*, Pro-gradu-tutkielma, Hoitotieteen laitos, Tampereen yliopisto, 2004.
- [9] Rauterberg M., *Cost Justifying Usability – State of the Art Overview 2003*, Eindhoven University of Technology, 2003.
- [10] Ravimo Petri, *KanTa-kokonaisuus*, Kansallinen eArkisto, eResepti ja eKatselu, Vaasan sairaanhoitopiiri, 2011.
- [11] Rintala Jyrki, *Pegasos: eArkisto*, CGI Group Inc., 2013. [http://www.cgi.fi/sites/default/files/files\\_fi/Brochures\\_publications/pegasos\\_earkisto.pdf](http://www.cgi.fi/sites/default/files/files_fi/Brochures_publications/pegasos_earkisto.pdf)
- [12] Rogers Yvonne and Sharp Helen, *Interaction Design*, Wiley, 2002.
- [13] Saari Sirpa, *Kynällä tai koneella – Sairaanhoitajien käsityksiä manuaalisesta tai tietokoneavusteisesta hoitotyön kirjaamisesta*, Pro gradu-tutkielma, Hoitotieteen laitos, Tampereen yliopisto, 1995.
- [14] Vänskä Jukka, Vainiomäki Suvi, Kaipio Johanna, Hyppönen Hannele, Reponen Jarmo ja Lääveri Tinja, *Potilastietojärjestelmät lääkärin työvälineenä 2014*, *Suomen Lääkärilehti* 49/2014. 3551-3358.
- [15] Vänskä Jukka, Vainiomäki Suvi, Kaipio Johanna, Hyppönen Hannele, Reponen Jarmo ja Lääveri Tinja, *Potilastietojärjestelmät tuotemerkeittain arvioituna vuonna 2014*, *Suomen Lääkärilehti* 49/2014, 1660-1667.

- [16] Walldén Sari, Miten tutkia potilastietojärjestelmien käytettävyyttä, *Finnish Journal of eHealth and eWelfare* 2, (3), 2010, 90-100.